# AN INTEGRATED APPROACH TO ANALYSE SHIPHANDLING EXPERTISE IN A FULL MISSION BRIDGE SIMULATOR


Capt. Luca Orlandi

MD in Psychology (Clinical and Community)
MD in Maritime and Nautical Sciences
BD in Electronic Engineering


Submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy
University of Tasmania, Australian Maritime College
March 2018

# ABSTRACT

The aim of this doctoral research is to increase and deepen the understanding of the concept of Expertise in marine pilotage. The research introduces a methodology that combines a suite of variables to evaluate levels of “shiphandling performance”. Ten Marine Pilots participated in the research using a Full Mission Bridge Simulator, however the number of pilots does not give a good understanding of the scale of the data collected. Four different manoeuvres performed by each participant were used as experimental conditions, resulting in 6-8 hours of continuously collected data for each pilot. Those manoeuvres were controlled on three factors: the level of difficulty (easy and difficult), familiarity with the port (homeport and foreign port) and manoeuvring phase (approach, swing and closing). The dependant variables included performance, physiological and behavioural measurements. Before performing the manoeuvres in the simulator, pilots were required to plan each vessel movement. The plan was required to identify the expected use of propulsion power, vessel positions and speeds. Through a face to face interview this information was translated in numerical values.

Dependent variables were obtained through a direct comparison between planned values and actual outcomes in the manoeuvres, as recorded by the simulator. Those variables were related to vessel’s XTD (Cross Track Distance), speed and propulsion power used. Statistically significant results were recorded in the factor phase, showing how the swing offered important challenges in the positioning of the vessel while rotating it. Pilots estimated the vessels speed differently in the two ports (underestimating the vessel’s speed in the foreign port) and for the two levels of difficulty (probably since dealing with two different type of propellers). Pilots underestimated the need of propulsion power in the swing phase for the more difficult manoeuvres. Such underestimation suggested further considerations with reference to safety, when manoeuvres are performed close to operative limits and maximum propulsion power available.

Physiological measurements, comprising EEG (power spectrum density distribution in the bands B1 and B2), ECG (heart rate and LH/HF index) and pupil dilation, were compared against self-reported measurements (Likert scale on seven levels and NASA TLX), to better appreciate the relationship between different techniques to assess mental workload. Results obtained from measuring ECG, EEG, and pupil dilation provided some indications that physiological variables correlated to scores obtained from self-assessment scales. Light correlations were identified between the self-assessment Likert scale, heart rate and pupil dilation. Increasing the level of difficulty induced a significant increment in the levels of responses, particularly in the HR.

The use of eye trackers facilitated the measurement pilots' gaze, and voice recordings identified variations in speech behaviour. When viewing these results through the lens of Smith and Hancock's Perceptual Cycle model of situational awareness, participant pilots were able to consistently direct their attention to specific and more relevant sources of information, depending on manoeuvring conditions. Significant results showed how gaze active search, was dependant and adaptive to the specific shiphandling tasks elicited by the manoeuvres. Eye trackers were able to document significant interactions between the subjects and their working environment, through the accounting of pilots' orders. Results showed how the frequency of those orders were significantly higher when more critical shiphandling conditions were experienced.

This research offers its contribution to the broader field of Expertise as applied to the Maritime Industry. It demonstrates how a set of empirical techniques can be used to assess a specific, yet multivariate skill – that of shiphandling. The study explores and aims to make a contribution in the validation and use of physiological measures as relatively unobtrusive proxies for mental workload. Results in this area can improve and increase the number of tools available to the shipping industry for the prevention of accidents. This research has moved our understanding of where the "red lines" of workload for marine pilots might be, levels beyond which the safety of the operations could be compromised. This study provides a tangible example of how more complex theoretical constructs such as situation awareness, can be unpacked in their constitutive elements (e.g., perception, attention). This research shows how the latest eye tracking technology can be profitably used to highlight differences and characteristics of those processes, and how those processes do change depending on specific contexts and goals. In sum, it depicts expertise as a combination of human and ship performance, embedded in a context of the difficulty of the task, the familiarity of the environment and the phase of the manoeuvre.

Defining a set of standard and unobtrusive measurements contributes to address a gap identified in Industry assessment standards and, more broadly to better understand and define the nature of Expertise in such specific environment. This research was conducted in a full mission bridge simulator, and the aim for the future would be to adopt similar techniques in the real world. Such analysis would be able to highlight areas of improvement where pilots' approach to manoeuvres could be discussed, reconsidered and modified.

## DECLARATION OF ORIGINALITY

This thesis contains no material which has been accepted for a degree or diploma by any tertiary institution and, to the best of the Candidate's knowledge and belief, no material previously published or written by another person except where due acknowledgement is made in the text of the thesis. The thesis does not contain any material that infringes copyright.

__________________ Date: 25/06/2017

Luca Orlandi

## STATEMENT OF ETHICAL CONDUCT

The research associated with this thesis abides by the *Australian Code for the Responsible Conduct in Research* (2007) and the *National Statement on Ethical Conduct in Human Research* (2007).

__________________ Date: 25/06/2017

Luca Orlandi

## STATEMENT REGARDING PUBLISHED WORK AND AUTHORITY OF ACCESS

The publishers of the papers comprising Chapter 6 hold the copyright for that content, and access to the material should be sought from the respective journals. The remaining non-published content of the thesis may be made available for loan, limited copying and communication in accordance with the *Copyright Act* (1968).

__________________ Date: 25/06/2017

Luca Orlandi

The manuscripts contained in this thesis comply with the University of Tasmania's *Authorship of Research Policy*. The following people and institutions contributed to the publication of work undertaken as part of the thesis:

**Luca Orlandi**, *Australian Maritime College, University of Tasmania*
= Candidate/Author 1

**Benjamin Brooks**, *Australian Maritime College, University of Tasmania*
= Author 2

**Marcus Bowles**, *Australian Maritime College, University of Tasmania*
= Author 3

**Joanne Wood***, Queensland University of Technology, Faculty of Health, School Optometry and Vision Science*
= Author 4

**Alexander Black***, Queensland University of Technology, Faculty of Health, School Optometry and Vision Science*
= Author 5

**Author details and their roles:**

**Paper 1** (located in Section 6.1)

Orlandi, L., Brooks, B., & Bowles, M. (2014). The development of a shiphandling assessment tool (SAT): A methodology and an integrated approach to assess manoeuvring expertise in a full mission bridge simulator. Paper presented at the 15th Annual general assembly International Association of Maritime Universities.

The Candidate is the main author and was primarily responsible for the conception, planning and execution of the work. Authors 2 contributed to the idea as well as its formalisation and refinement. The Candidate's contribution is approximately 80%.

**Paper 2** (located in Section 6.2)

Orlandi, L., Brooks, B., & Bowles, M. (2015). A Comparison of Marine Pilots' Planning and Manoeuvring Skills: Uncovering Mental Models to Assess Shiphandling and Explore Expertise. Journal of Navigation, 1-18.

The Candidate is the main author and was primarily responsible for the conception, planning and execution of the work. Authors 2 contributed to the idea as well as its formalisation and refinement. The Candidate's contribution is approximately 80%.

**Paper 3** (located in Section 6.3)

Orlandi, L., & Brooks, B. (2018). Measuring mental workload and physiological reactions in marine pilots: building bridges towards redlines of performance. Applied Ergonomics, Vol. 69, 74-92.

The Candidate is the main author and was primarily responsible for the conception, planning and execution of the work. Authors 2 contributed to the idea as well as its formalisation and refinement. The Candidate's contribution is approximately 80%.

**Paper 4** (located in Section 6.4)

Orlandi, L., Brooks, B., Wood, J., & Black, A. (201x). Interpreting changes in marine pilots' perceptual cycle through gaze detection patterns. Ergonomics (under review).

The Candidate is the main author and was primarily responsible for the conception, planning and execution of the work. Authors 2, 4 and 5 contributed to the idea as well as its formalisation and refinement. The Candidate's contribution is approximately 80%.

We, the undersigned, agree with the above-stated proportion of work undertaken for each of the published, accepted or submitted peer-reviewed manuscripts contributing to this thesis.

Candidate/Author 1: ____________________ Date: 25/06/2017

Luca Orlandi

Author 2: ____________________ Date: 25/06/2017

Benjamin Brooks

Author 3: ____________________ Date: 25/06/2017

Marcus Bowles

Author 4: ____________________ Date: 25/06/2017

Joanne Wood

Author 5: ____________________ Date: 25/06/2017

Alexander Black

# STATEMENT OF ETHICAL CONDUCT

The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University.

____________________ Date: 25/06/2017

Luca Orlandi

# ACKNOWLEDGEMENTS

..And here we are, exactly 6 years after that 24th of June 2011, for that meeting at the Launceston Airport with Dr. Benjamin Brooks…!

At that time an ash cloud coming from a volcanic eruption in South America was paralysing the air traffic in the whole South Pacific region. Dr. Brooks and myself were literally trying to cross each other in the sky. I was coming from Europe and aiming to catch a ship in New Zealand and Dr. Brooks was back from the mainland, keen to immerse himself again in the wonders of the Tasmanian winter.
That's how it all started.. Ash clouds in the sky, Tasmanian winter howling out of the airport windows.. the signs were all there.. I should have known better that it wasn't going to be easy!
..but what a kaleidoscopic adventure was just about to start!

Since then, a few more other planes, other crosses in the sky (or wherever else Dr Brooks allowed me to chase him down!) until today..

the day of my PhD thesis submission!

..Guess that's a bit of a succinct summary of the last few years.. but this is not an autobiography, is it?!
In reality many more other things have happened, and they wouldn't have ever been possible without the amazing support received from every person that helped me out along the way of this incredible journey, including those that were there at the beginning and decided otherwise along the way.

First of all, I would like to thank all the pilots that participated to this research. I cannot mention your names, but I really want to express my gratitude to you all. Without your enthusiasm, your patience, your availability to bear the annoying devices while playing

with me in the simulator, this work would have never been possible. Thanks for your participation and your time. I was able to learn so much and now that I am piloting ships myself, I can only appreciate more your dedication and professionalism. I owe you a big one!
Thank you Capt. Peter Liley and thanks to the Australasian Marine Pilot Association (thank you Helen Gray!), for supporting my studies and pushing my crazy ideas where they would have never been heard otherwise. Thank you Capt. Rob Buck for continuing to think that this was a good idea! …Peter, I hope you didn't get too much of a beating for that! Peter, Patricia and your beautiful family, thank you all, for being my wonderful Australian family! Patricia, thanks again for the hints on the eye trackers …what a tip!!

Thank you Capt. Peter Listrup, for hosting me day (and night) at the Maritime Safety Queensland Smartship Simulator in Brisbane. Peter, I am deeply grateful to you. Without your availability, I wouldn't have had the opportunity to learn and to see so much. I wouldn't have had the honour and the privilege to be introduced to Capt. Ravi Nijjer, Capt. John Watkinson.. and so many other incredible people!
My gratitude certainly goes to the team that was at Smartship Simulator at the time. Thanks Capt. Jeremy Brew and Amanda, for letting me play with your beloved super dooper vinyl turn table during weekends, without showing fear! ..Olly and I had a great time! Thank you Hugh Ferrar, for your patience and your stoic aplomb when it was time to collect the pieces of my IT damages.. thanks for sending me on a sailboat in the unconcealed attempt to keep me away from your PCs! Ms. Tori Bradley, I thank you for everything you have done for me, for every little daily gesture, including saving some lunch for a poor student! ..if it wasn't for you certain beautiful things wouldn't have never happened …right?! ;)
Thanks Capt. Brenton Winn, and thanks to Mrs. Claudette Perrett, Mrs Michelle Weinberg and the amazing people that I met at BMP. Thanks for your help and your support! This thesis wouldn't be here if that door wasn't open that day!
Thanks Capt. Hans Hederstrom and the team at C-Smart, for hosting and listening, when I still had to figure out what I was saying (..it hasn't changed much!).

Thanks Dr. Elkana Ngwenya, Prof. Irene Penesis, Prof Alberto Zuliani, for throwing me a life ring of wisdom at the time I was drowning in the shark infested ocean of statistics! Thanks Capt. Richard Dunham, for letting me play at being the teacher for some time.. bad cop - good cop.. that was fun! …reality is that there is still so much for me to learn! Thanks Dr. Marcus Bowles for your tips and suggestions. I am looking forward to having the opportunity to catch up with you again. Thanks to all my friends in Launceston and around Australia - Georgia, Cecile, Mahdi, Ahmed, Rahman and all those brave hearts that decided to venture in the conquest of a PhD… I am actually the one still lagging behind!!!

An enormous thank you to the team at the Queensland University of Technology. Thanks Prof. Joanne Wood and Dr. Alex Black from the bottom of my heart for your availability and generosity ..I still struggle to understand how on earth, you allowed someone looking like a terrorist like me, to knock on your door, kindly ask to borrow your tens of thousands of dollars' worth of equipment and actually go away with it, under the sole solemn promise that I returned it.. (I am struggling even more to believe that, eventually, I brought the equipment back!). Thank you Dr. David Alonso-Caneiro for your providential crash courses in Matlab.. what a great toy you showed to me!

Bel, thank you, for sharing with me every step up and down those rolling hills, and then the mountains, and then the beaches, and then the fields, and then…! ☺

..guess it is your turn Dr. Brooks!

Thank you for your help, patience, guidance, headaches, moments of rage and moments of joy…!

It has certainly been the incredible experience that we anticipated (..and more!)!

..It might still be winter, but I see the sky now!

☺

Luca

## TABLE OF CONTENTS

LIST OF FIGURES

# LIST OF TABLES

# LIST AND STATUS OF PUBLICATIONS

This research includes four papers that have been published or are under review. An unpublished version (Note 1: Additional Version) of paper IV was provided as paper IV – AV. These papers have been reproduced in Chapter 6.

| Paper | Nature | Status | Title | Publication channel | Publication details | Full-length, double-blind review |
|---|---|---|---|---|---|---|
| **I** | Conceptual paper | Published (2014) | The development of a shiphandling assessment tool (SAT): A methodology and an integrated approach to assess manoeuvring expertise in a full mission bridge simulator. | *Paper presented at the 15th Annual general assembly International Association of Maritime Universities*. | ISBN number 098063914X Proceedings pp. 131-140 | Yes |
| **II** | Research paper | Published (2015) | A Comparison of Marine Pilots' Planning and Manoeuvring Skills: Uncovering Mental Models to Assess Shiphandling and Explore Expertise. | *Journal of Navigation* | ISSN number 1469-7785 Vol. 68, Issue 5, pp. 897-914 | Yes |
| **III** | Research paper | Published (2018) | Measuring mental workload and physiological reactions in marine pilots: building bridges towards redlines of performance. | *Applied Ergonomics* | ISSN number 0003-6870 Vol. 69, pp. 74-92 | Yes |
| **IV** | Research paper | Submitted | Interpreting changes in marine pilots' perceptual cycle through gaze detection patterns. | *Ergonomics* | Not yet applicable | Yes |
| **IV – AV(1)** | Research paper | Unpublished | Interpreting changes in marine pilots' perceptual cycle through gaze detection and speech patterns. | *(1) Additional Version Unpublished* | Not applicable | No |

# 1. INTRODUCTION

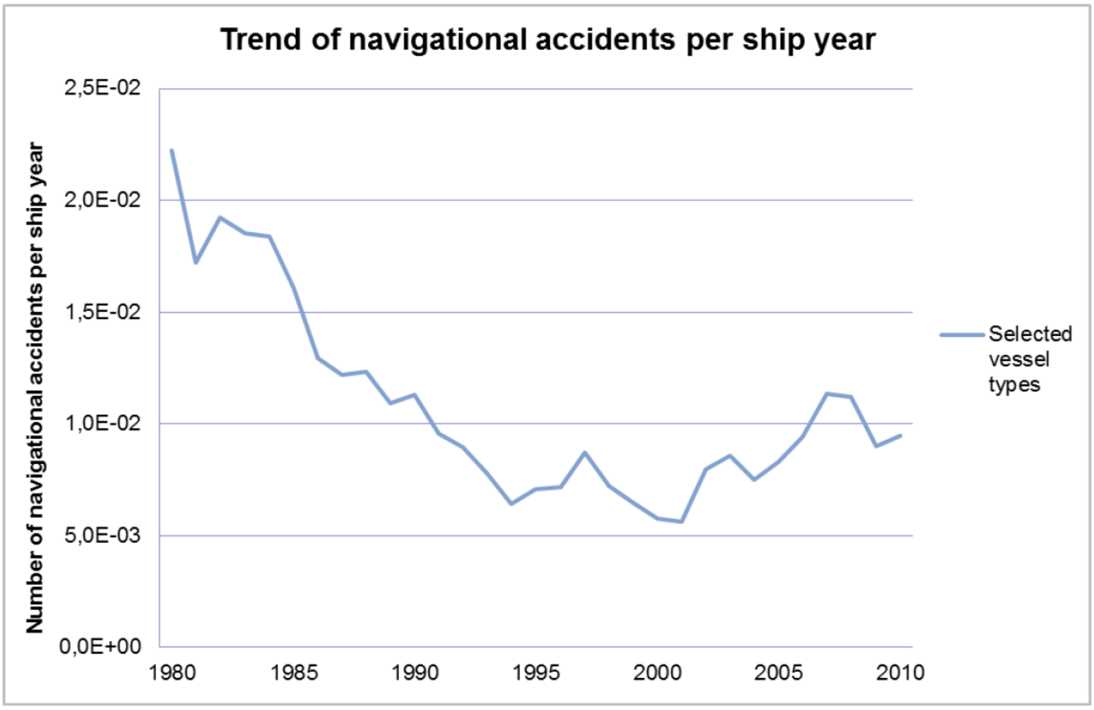
Chapter 1 introduces the general topic and main concepts of this research. It also provides definitions of the key terms, to identify the research gaps, and to formulate a set of research questions designed to guide the entire doctoral project.

## 1.1. Research background

Shipping is the predominant transport mode for the movement of freight in the world. Commercial vessels carry around 90% of the world trade. The maritime industry generates an annual income of over half a trillion US dollars in freight rates, with a worldwide population serving on internationally trading merchant ships in the order of 1,187,000 seafarers (ICS, 2016). Seafaring, compared to any other sector, can be considered perhaps the most globalised labour market in the world (Alderton, 2004), with labour force drawn from an unrivalled number of different countries.

Even though some authors consider the shipping industry having a reasonable safety record (Hetherington, Flin, & Mearns, 2006), maritime incidents have a high potential for catastrophes (Faghih-Roohi, Xie, & Ng, 2014). A study conducted in the US comparing different transport modalities, reports that the workplace fatality rate per 1000 employees in the maritime transportation (0.24) is four times as high as the one in the air transportation (0.06) (Savage, 2013). In 2013, a report from the IMO (international Maritime Organization) Correspondence Group on E-navigation provided statistics based on the IHS Fairplay casualty database to support this initiative (HIS is considered perhaps the most complete and reliable maritime data source in the world). Despite a previous decreasing trend recorded from 1980 until 2000, this report identified that the total number of navigational accidents on cargo, passengers and offshore ships increased between 2001 and 2010 (from less than 400 in 2001 to more than 700 in 2010). The report also identified how the number of accidents per ship increased from 0.5% in 2001 to 1% in 2010 (see Figure 1). Of the total number of accidents considered, 22% were groundings, 22% were collisions and the rest were classified as other types. The same document reported that around 65% of collisions and groundings were caused by human error, while only 18% were caused by technical failure and 17% by external factors. Some of the human error causes identified by the report were: inattention (28%), poor judgement of ship movement (17%), work overload (13%), poor judgement of other factors (12%), inadequate planning of voyage (9%), Inadequate use of navigational aids (3%), lack of skill or knowledge (3%) etc.. These are all topics that will be considered in the present thesis (IMO, 2013). Other studies identified other systemic factors as contributing to maritime accidents, such as the social organization of the personnel on board, economic pressure, ‘hidden’ ownership structures, and challenges associated with international regulation (Perrow, 2011). At an individual level, long contracts, limited sleep opportunities between shifts and short turn-around times can create fatigue, stress and work pressure, exacerbated by a lack of training (McNamara, Collins, & Mathews, 2000). Although the maritime industry established a set of internationally recognised and accepted standards for seafarer training and certification (STCW) (IMO, 1978), many differences can be found in the modalities through which such certifications are issued, endorsed and renewed worldwide.

Figure 1. Trend of Navigational accidents per ship per Year (IMO, 2013)




Although this thesis is focused on the quantification and qualification of shiphandling expertise, and several human factors issues associated with that competency, the implications sit squarely in the domain of seafarer training. One of the increasingly important options in seafarer training is computer based assessments (CBA). Today, CBA, as adopted by the maritime industry tends to rely on multiple choice questionnaires or simple desktop simulations. Those tools are mainly used to provide a relatively unsophisticated screening during the initial stages of personnel selections or for Certificate of Competency issuing purposes (Gekara, Bloor, & Sampson, 2011). Regardless of the type of approach or tools adopted by each single Nation, a considerable variation in assessing standards (Ghosh, Bowles, Ranmuthugala, & Brooks, 2014) within each Country exists. This makes it difficult for employers to rely upon seafarer licences as an indication of seafarer competence, skill, or knowledge (H. Sampson, Bloor, & Gekara, 2011).

Several challenges remain to seafarer education and training (Helen Sampson, 2004; Helen Sampson & Bloor, 2007). One of those challenges is the need to identify a commonly accepted and standardized way to assess shiphandling competency and performance (Hsu, 2015; Kobayashi, 2005), also in the context of more complex organizations (Bruzzone, Longo, Nicoletti, & Diaz, 2012). As previously mentioned, merchant vessels can reach displacements of hundreds of thousands of tons and a simple accident has the potential for disastrous consequences. Assessing shiphandling competency simply through oral examinations or multiple choices tests, is inadequate and high risk. Even simulator-based assessment tends to involve subjectivity and lacks rigour. The ability to perform workplace tasks, assessed through methods that resemble professional scenarios, becomes an important element. Such performance-based assessments applied in real-world contexts have often been described as *'authentic assessments'* (Ghosh, Bowles, Ranmuthugala, & Brooks, 2016).

This is well known by Marine Pilot Companies that have to make considerable investments before the necessary training and assessment period for a newly recruited trainee pilot can be considered completed and satisfactory, according to port tailored standards and practises. A Canadian report (CMPA, 2017) explains how the use of Marine Pilots is one of the most effective measures that are adopted in the shipping industry to mitigate accidents. The reports highlights how piloted ships are able to have their risk reduced 44 times compared to not piloted ships (from 0.094 to 0.0021 probability of accident per vessel). The risk of collision and grounding drops 12 times more if a piloted vessel has also tugs in assistance (from 0.0021 to 0.00018 probability of accident per vessel). Increasing the level of training of personnel seems to be one of the keys to success and, for this purpose, simulators offer a potential solution.

Simulators have significantly improved in the last few years and they are at a point where they have proven their value in different fields of application within the transport industry, for example to simulate logistic dynamics and volumes before a port is even built (Ryan, 1998), to improve training for train drivers (Naweed & Balakrishnan, 2012) as well as for airline pilots (Dahlström, 2008). The aviation industry has relied on simulation perhaps more than any other safety critical industry. While simulators are still used for stick-and-rudder and instrument training, today they are also part of practically all aspects of aviation training (Salas, Bowers, & Rhodenizer, 1998) and the new multi-crew pilot licence (MPL) rests almost entirely on simulated flight training. Simulators, being able to allocate latest trends in shipping constructions, become fundamental for studies on shiphandling and port development (Perkovic, Brcko, Luin, & Vidmar, 2016). This investment in simulation reflects an industry-wide confidence that it can save time, money and lives (Bürki-Cohen, Soja, & Longridge, 1998) and, in addition, it can provide effective training (Hontvedt & Arnseth, 2013), developing skills and knowledge that are transferable to any target situation (Dahlstrom, Dekker, Van Winsen, & Nyce, 2009).

In this context, the aim of this doctoral research was to increase and deepen the understanding of Shiphandling Expertise, with the use of a full mission bridge simulator. The intent was to introduce and test a methodology that, taking into account several empirically measurable variables, could evaluate seafarers' levels of such expertise.

## 1.2. Research definitions

This research was conducted within the context of maritime pilotage. Marine Pilots are ship's captains that are specifically trained and certified to manoeuvre vessels within critical coastal and port waters. They embark a ship outside port waters and then work with the bridge team to navigate the ship to berth. While the ships' Captain still retain the full charge or responsibility of the safety of the vessel, pilots generally take the "conduct", manoeuvring the ship in enclosed and or critical waters until a safer position is reached or the vessel is alongside the assigned mooring. In Australia pilots are provided by several companies or organizations (private, government or port authority), that are responsible for providing pilotage services in a particular port or district.

In this study pilots are referred as "experts". Doing so, experts are defined as those who have acquired noticeable skills or knowledge of a particular subject, through training and practical experience, capable of recalling complex, task specific patterns gaining access to the right information (Scardamalia & Bereiter, 1991). As experts, pilots are expected to be specialists having specialised knowledge (Mieg, 2001), able to restructure, reorganize, and refine their representation of knowledge, applying it more efficiently into their environment, with their expertise being the result of a complex adaptations of mind and body, exploiting substantial self-monitoring and control mechanisms, to the tasks and goals imposed to them by the environment (K. A. Ericsson & Lehmann, 1996). Their actions were expected to be smoother and more efficient, and performance to be achieved with a minimal effort, running essentially automatically, with minimal cognitive control (Posner & Snyder, 2004). As experts, pilots are expected be able to run more processes in parallel, thanks to the reduction in the cognitive demand due to automaticity (Shiffrin & Schneider, 1977).

Those assumptions were adopted, when inviting pilots as participants in this research. The aim was to identify those specific elements and characteristics that would have helped us to quantify their expertise. Since pilotage involves a complex interaction between the pilot and a bridge team, tug masters, a vessel traffic service and electronic equipment, the study herein presented tries to unpack such complexity, considering several theoretical constructs.

Figure 2. Model of Shiphandling Expertise adopted in the research

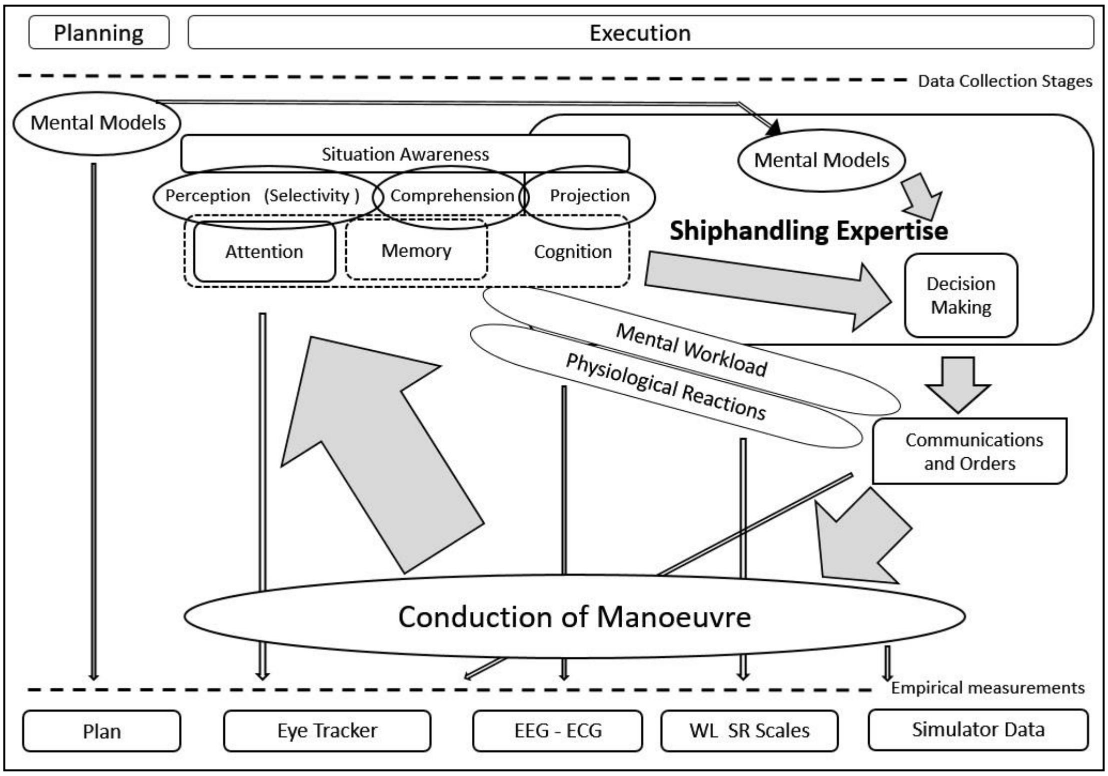



Figure 1 briefly sketches how in the initial developmental phases of this research, the aim was to unpack Pilots' Expertise, capturing key constructs and using empirical measurements. The figure also links key aspects of expertise with human factors issues.

The expectation was that pilots, as experts, when involved in planning activities, would be competent to forecast future developments with a high degree of precision (M. T. H. Chi, Glaser, & Rees, 1981). To do this, they should be able to evaluate initial conditions and to structure them into more accurate and realistic mental models. The analysis of those mental models, captured by the plans, was the starting point of this study.

Mental models have been succinctly defined as "mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future states" (Rouse & Morris, 1986) p.351). In this research, the manoeuvres proposed were used to be able to obtain an understanding of the pilots' plan for the manoeuvre. By inference, this provided an indication of their mental model for the manoeuvre. For research purposes the plan was translated into a numerical format for comparison with performance in the simulator. Those models, once obtained, showed pilots' capability to understand the specificities and the implications of the required manoeuvres. The expectation was that those mental models were able to explain and direct pilots' decision making, working as a guide (Mumford et al., 2012) or as a map (Fiol & Huff, 1992). The expectation was also to witness pilots' attention to be directed to the environment accordingly to the mental model adopted, and to perform an efficient filtering of available stimuli. Such filtering activity, despite carrying the risk of omitting relevant data, becomes necessary whenever large amount of information is available.

In the proposed model, perceived elements are expected to be integrated in a meaningful ensemble and to be compared to related contents retained in memory structures. Elements gathered from reality would have confirmed or not if the mental model adopted was substantially 'correct'. "Testing against reality" is a clear reference to a continuously maintained state of situational awareness (M. R. Endsley, 1988) through the exploitation of the underlying cognitive processes of perception, comprehension and projection.

It is arguable that the comparison between the managed situation awareness and the embedded mental model, triggered pilots' decisions regarding actions deemed necessary. In the context of this research, decision making specifically referred to the naturalistic paradigm where expertise is evaluated in its naturalistic context (Bornstein, Christine Emler, & Chapman, 1999; M. S. Cohen, 1993; Keren, 1987; G Klein, Shafer, & Ross, 2006; Shanteau, 1989; J. F. Smith & Kida, 1991).

The conduction of the manoeuvre was the practical translation of pilots' decisions into competent behaviours. This thesis argues that those behaviours, or interactions with the environment, can be identified by specific gaze searches and effective communications and orders. This was considered an iterative process. The iteration is highlighted in Figure 2 through thicker arrows. Once pilots made a decision, they acted on the environment through communication and orders. They then assessed the outcome, directing their attention on specific and meaningful elements. This active gaze searches fed the cognitive processes of perceiving, understanding and projecting into the future, which are fundamental to gain and maintain situation awareness. Comparing the current situation with the goal, as embedded in their mental model, pilots had then the opportunity to make a new decision. Should further actions be required to address discrepancies between what aimed and what experienced, pilots could redirect their actions through new communications and orders. This entire process would have repeated itself throughout the conduction of the manoeuvre until its completion.

Every step of the described process and its constant iteration places a certain “burden” on the shoulders of the pilots involved. This “burden” might be otherwise described as mental workload. Despite the interest in the topic for the last few decades (Huey & Wickens, 1993), there is still no clearly defined and universally accepted definition of mental workload (Cain, 2007). Workload, as a mental construct, is considered a variable (Gopher & Donchin, 1986), dependant on the mental demands imposed by the different required tasks and the operator’s experience in that particular context, and constrained by the ‘resources’ available to manage it. Workload is thought to be multidimensional and multifaceted, difficult to be uniquely defined. Since workload cannot be directly observed, overt measurements of psychological and physiological variables were gathered and used for inference (Casali & Wierwille, 1984).

At the bottom of Figure 2 it is possible to identify the relevant group of variables that were identified in this research for each construct, and that will be extensively described in subsequent paragraphs. Using the model of expertise identified, the research was then oriented to define key research questions and to identify and test suitable methods and measurement techniques. These research questions are described in the next section of this thesis.

## 1.3. Research gaps and questions

After having introduced the main concepts of this research in paragraph 1.1. and research definitions in paragraph 1.2., this section identifies the three research gaps that have guided the research presented in this thesis.

### 1.3.1. Evaluating pilots' performance

As mentioned in paragraph 1.1., it is asserted that the empirical assessment of Pilots' shiphandling competencies would provide significant benefit to the Maritime Industry. The first research gap relates to the fact that, even though a competency based system is in place and it is internationally adopted (IMO, 1978), standards and procedures to specifically and empirically evaluate shiphandling competency have yet to be established and endorsed. Each shipping or pilot company, each port authority refers and relates to own criteria and admission tests. Those tests are not generally accepted or standardised, nor are the results easily transferrable to other contexts other than the one within which they were developed.

*1st Research GAP: Need to identify standards to evaluate shiphandling performance.*

Hence, the first research questions are as follows:

1. *How would we quantify both the human and ship performance in order to derive an assessment of shiphandling expertise?*
2. *How can these variables be sufficiently general to be adopted in different shiphandling conditions, such as different ports, environmental conditions, etc..?*

### 1.3.2. Evaluating pilot's workload

Shiphandling is a complex activity that requires the seafarer to manage several tasks at the same time. To conduct a vessel during a berthing manoeuvre is an operation that involves ships' crews, tug skippers, shore parties, Vessel Traffic Management stations and last, but not least, other vessels present in the area. The presence and the management of all those elements require pilots' attention and have an impact on pilots' workload.

Mental workload is a multidisciplinary concept (M. S. Young, Brookhuis, Wickens, & Hancock, 2015) and has long been recognized as an important element of human performance (Eggemeier, Wilson, Kramer, & Damos, 1991; Parasuraman, Sheridan, & Wickens, 2008), particularly important in high risk environments (Jou, Yenn, Lin, Yang, &

Chiang, 2009) and those demanding high levels of reliability (Carswell, Clarke, & Seales, 2005; Yurko, Scerbo, Prabhu, Acker, & Stefanidis, 2010). Mental workload varies around a combination of task demands and resources that a particular individual has available (Noyes, Garland, & Robbins, 2004; M. Young & Stanton, 2005). From this “resource-based view”, mental workload can be seen as the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support and experience.

For the purposes of this study, mental workload followed the definition of subjects’ direct estimate or comparative judgment of mental or cognitive effort experienced at a given moment (Luximon & Goonetilleke, 2001). Debate continues in the maritime domain around issues related to workload with reference to increasing automation, use of technology, and even remote pilotage (Brooks, Coltman, & Yang, 2016). Presently, though, little is known about the extent to which ship manoeuvres create workload and, more importantly, if excessive workload might breach acceptable levels. Even though qualitative studies have been conducted (M. H. Lützhöft & Nyce, 2006), as yet it is not possible to define an acceptable level of workload and the physiological implications for pilot’s workplace health and safety.

*2nd Research Gap: A need to find reliable measures of workload helping to highlight when workload might exceed resources available to perform the task (both internal and external).*

The second group of research questions are reported below:

3. *How can we empirically and unobtrusively measure workload during shiphandling manoeuvres?*
4. *Does the level of workload experienced have a relationship with manoeuvring conditions and outcomes?*

### 1.3.3. Evaluating pilots’ situation awareness

Pilots are the interface between the port and the ship that intends to call in that port. Their role is primarily to ensure the safety of the operation, meaning the safety of all the people and the means involved. Piloting in a port is mainly taught by more experienced pilots during an initial training period and then it is endlessly refined through the years of practice that will follow. But what is it exactly that pilots learn and experience, and what are those elements that are taken into consideration while piloting? Would the consideration of certain elements be dependent on the particular manoeuvring context? These are some of the many questions that still remain to be answered.

*3rd Research Gap: Need to find reliable measures for pilots' situation awareness (collection of information and direction of actions).*

Hence, the final research questions investigated in this research are:

5. *How can pilots' efforts to gain and maintain situation awareness be empirically and unobtrusively measured during shiphandling manoeuvres?*
6. *How can we identify pilots' specific behaviours that will vary depending on manoeuvring conditions?*

All of these research questions combine to provide an understanding of the broader topic of marine pilot expertise. This thesis therefore makes a significant contribution by assessing aspects of expertise and linking them with human factors concepts through empirical analysis in a simulated environment.

### 1.3.4. Linking research questions to published and submitted work

This thesis forms a capstone to published and submitted studies conducted within the PhD project from March 2012 to June 2017. In particular, the doctoral work contains four fully refereed papers prepared during the candidature period. An additional version (unpublished) of one of the papers was added, since it was reporting results not included in the submitted version. The research presented in these papers as well as their respective findings are synthesised in Section 4.1, and the actual publications are reproduced in Chapter 6. Two of the papers have been published and two are under review.

This capstone thesis is designed to demonstrate that the above-mentioned papers constitute essential parts of a coherent and integral body of work related to a single research project and a set of related questions. As indicated in Table 1, the four reviewed papers included (as well as the additional version of paper IV) in this thesis are linked thematically to the research gaps and questions presented in Section 1.3..

Table 1. Linking the papers to the research gaps and the research questions

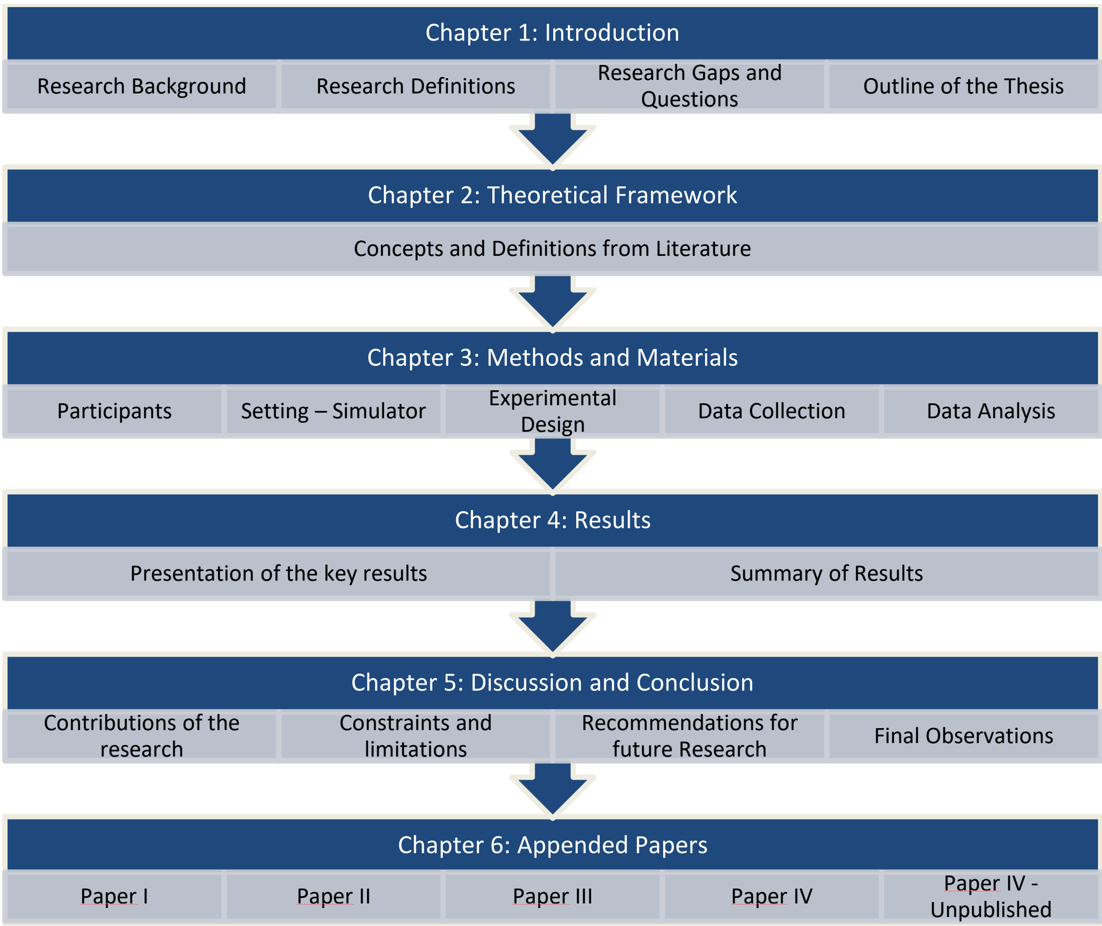
| Research questions | | Research gaps addressed | Related papers |
|---|---|---|---|
| 1. | How would we quantify both the human and ship performance in order to derive an assessment of ship-handling expertise? | Research Gap #1 | Paper II |
| 2. | How can these variables be sufficiently general to be adopted in different shiphandling conditions, such as different ports, environmental conditions, etc..? | Research Gap #1 | Paper II |
| 3. | How can we empirically and unobtrusively measure workload during shiphandling manoeuvres? | Research gap #2 | Paper III |
| 4. | Does the level of workload experienced have a relationship with manoeuvring conditions and outcomes? | Research gap #2 | Paper III |
| 5. | How can pilots' efforts to gain and maintain situation awareness be empirically and unobtrusively measured during shiphandling manoeuvres? | Research gap #3 | Paper IV |
| 6. | How can we identify pilots' specific behaviours that will vary depending manoeuvring conditions? | Research gap #3 | Paper IV |

## 1.4. Outline of the thesis

This research increases and deepens the understanding of Shiphandling Expertise, specifically in the context of Maritime Pilotage. Through the different papers presented, it demonstrates how a multivariate concept such as expertise can be unpacked in fundamental components and processes. This research will also demonstrate how these fundamental components can be analysed using empirically measurable variables.

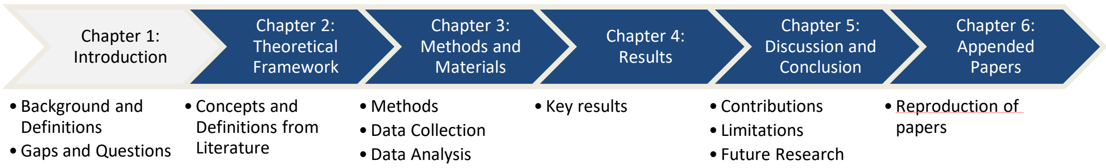
Figure 2 represents graphically the structure of this thesis. After having introduced the general topic of this study, identified research gaps, and formulated a set of associated questions in Chapter 1, the remainder of this thesis is organised as follows: Chapter 2 discusses the theoretical framework of this study and, in particular, the theoretical constructs that underpin the research. Chapter 3 provides details on the methods and materials used in this research and, specifically, on the experimental design chosen for the study (Section 3.3). The approach selected to collect data (Section 3.4) and to analyse it (Section 3.5) is also discussed. Chapter 4 focuses on the results published (and under review) in the framework of this research. The findings reported in the respective publications are summed up in Section 4.1, and a summary of these results is presented in Section 4.2. The results are discussed in Chapter 5. In particular, the theoretical and practical implications of the research are presented in Section 5.1. The constraints and limitations of this study are also addressed (Section 5.2), future research avenues are recommended (Section 5.3), and final observations are provided (Section 5.4). Section 6 reproduces each of the papers included in this thesis.

Figure 3. Thesis outline




As illustrated below, the next chapter will focus on the theoretical framework of this study. It will present the theories that underpin the research and its conceptualisation. It will also explain why empirical and unobtrusive measures are necessary to investigate and capture the multiple facets of the concept of shiphandling expertise.

Figure 4. Progress Tracker – Moving to Chapter 2



# 2. Theoretical framework

This research draws on existing theoretical foundations, concept and constructs. Chapter 2 presents these concepts as their respective contributions to the research. It also explains why it is necessary to go beyond these theories and to adopt an empirical approach, able to translate those concepts into objective measures.

## 2.1. Expertise

The literature about expertise is complex and vast. Possibly the first studies about Expertise are ascribed to Sir Francis Galton (1869), where he was attempting to identify a common set of hereditary causes to "Excellence" (K. Anders Ericsson, Krampe, & Tesch-Römer, 1993). Subsequently a plethora of researchers have investigated different phenomena related to Expertise in cognition, such as memory limitations and reasoning biases in knowledge representation. Some authors define Expertise as the considerable amount of skills, knowledge, and mechanisms that monitor and control cognitive processes to perform a delimited set of tasks efficiently and effectively (Feltovich, Prietula, & Ericsson, 2006). Following this definition, the "Experts" are then those who acquired noticeable skills or knowledge of a particular subject, through training and practical experience (Babcock, 1976). According to Scardamalia and Bereiter (1991) Experts are those individuals capable of recalling complex task specific patterns and of gaining access just to the right information with ease. Mieg ((2001), p. 4) sees "experts as specialists having specialised knowledge", introducing the idea that an "expert-by-experience must be an expert in the field", while an "expert-by-knowledge can be an expert about the field, while lacking personal experience in the field". Nonetheless Experts are able to restructure, reorganize, and refine their representation of knowledge, applying it more efficiently into their environment (K. A. Ericsson & Lehmann, 1996). Experts know more, know better (or differently) and can do significantly better than others. Their Expertise is the result of a complex adaptations of mind and body, exploiting substantial self-monitoring and control mechanisms, to the tasks and goals imposed to them by the environment. Table 2 is adapted from Farrington-Darby and Wilson (2006) and summarizes several characteristics of expertise from three different reviews (Cellier, Eyrolle, & Marine, 1997; M. T. H. Chi, Glaser, & Farr, 1988; Shanteau, 1992):

Table 2. Psychological characteristics and strategies of experts

| **Characteristics (Shanteau 1992)** | **Characteristics (Glaser and Chi 1988)** | **Characteristics (Cellier et al. 1997)** |
|---|---|---|
| *Extensive and up to date content knowledge.*<br><br>*Highly developed perceptual-Attentional abilities.*<br><br>*Sense of what is relevant when making decisions.*<br><br>*Ability to simplify complex problems.*<br><br>*Ability to communicate.*<br><br>*Handle adversity better.*<br><br>*Experts are better at identifying and adapting to exceptions.*<br><br>*Self confidence in decision making.*<br><br>*Adapt decision strategies to changing task conditions .*<br><br>*Strong sense of responsibility and willingness to stand behind their recommendations.*<br><br>*Use of strategies.*<br><br>*Willingness to make continuous adjustments in initial decisions.*<br><br>*Experts get help from others to make better decisions.*<br><br>*Experts often make use of formal or informal decision aids.*<br><br>*Experts make small errors they try to avoid making large mistakes.*<br><br>*They operate as though coming close is generally good enough.*<br><br>*Experts follow some sort of divide and conquer strategy.*<br><br>*Break problems down.* | *Experts excel mainly in their own domain.*<br><br>*Experts perceive large meaningful patterns in their domain.*<br><br>*Experts are fast (faster than novices at performing the skills of their domain) and they quickly solve problems with little error.*<br><br>*Experts have superior short term and long term memory.*<br><br>*Experts see and represent a problem in their own domain at a deeper (more Principled) level than novices; novices tend to represent a problem at a superficial level.*<br><br>*Experts spend a great deal of time analysing a problem qualitatively.*<br><br>*Experts have strong self monitoring skills.* | *Experts have greater skill in producing inferences when monitoring the values of variables, in using covert variables in building up a representation during diagnosis and in using inference strategies during the executive control of processing and task completion. In other words they can see the meaning behind the information provided and the implications of their decisions and actions.*<br><br>*Experts have greater skill in anticipating. They process cues preventatively rather than reactively during disturbances. They make better predictions of process evolution and changes in a system.*<br><br>*Experts have a more global and functional view of a situation and take a wider range of data into account in diagnosis. They operate through a limited number of assumptions that include the most relevant information, and account for possible side or spin-off effects through inference and anticipation.*<br><br>*Experts encode new information more quickly and completely.*<br><br>*Experts have more complete representations of the task domains.*<br><br>*Experts are considered to have a richer repertoire of strategies and appropriate mechanisms for assessing and applying strategies and the appropriate organisation of knowledge.* |

In the following paragraphs several theoretical concepts and how they relate to Expertise, will be introduced. How this research refers and adopts those concepts is graphically depicted in Figure 2 in section 1.2.

### 2.1.1. Expertise and Domain Specificity

This doctoral research was clearly conducted in a specific domain: Shiphandling. The association between Expertise and domain specificity has been well documented. Ericsson, Charness, et al. (2006), remind that it is quite uncommon for people to reach an elite level in more than a single domain of activity (K. A. Ericsson & Lehmann, 1996). One way that researchers use to understand the essence of expert performance in a specific domain, is to standardize representative tasks that can be presented to a group and then identify those skills and results that best discriminate experts from novices.

Asking Experts to repeatedly perform these types of tasks, allow experimenters to identify those complex mechanisms that mediate expert's superior performance (K. Anders Ericsson, 2006a). Where perceptual-motor skills are deeply involved (P.M. Fitts & Posner, 1967), such as in sports (M. Williams, 2004) or in music (Gruber, Degner, & Lehmann, 2004), domain specificity becomes even more apparent. It seems that there is virtually no limit to the level of specificity that expertise can reach within a particular field. In medicine, for example, different authors (Barrows, 1978; Elstein, Shulman, & Sprafka, 1978) showed that the same physician can demonstrate widely different levels of competence, depending on his or her particular experiential history. Experts in chemistry performed very much like novices, in tasks involving political science (Voss, Greene, Post, & Penner, 1983). In mental spatial manipulation of geometrical volumes, a study was able to show how 53 Tetris players outperformed 45 non Tetris players on a first experiment about mental rotation of shapes that were either identical to or very similar to Tetris shapes, but not on other tests of spatial ability (Sims & Mayer, 2002).

Chartrand, Peretz et al (2008), investigated domain specificity in processing and neural correlates of the human voice compared to face perception. Experts are not only differentiated from novices in the amount of owned knowledge, but also in the capability to learn new material, but only if the material is relevant to their field of expertise (Chiesi, Spilich, & Voss, 1979; Spilich, Vesonder, Chiesi, & Voss, 1979). The ability to learn from texts about soccer and baseball, for example, was more dependent on the learner specific expertise in that sport, than on his verbal IQ (Hambrick, 2002). Similar results were obtained in computer programming (Adelson, 1981; McKeithen, Reitman, Rueter, & Hirtle, 1981; S. Sonnentag, Niessen, & Volmer, 2009). Research has shown that Expertise, as well as general abilities such as learning, reasoning, problem solving, and concept formation, correspond to capacities and abilities that cannot be studied independently of the content domains. Garrett, Caldwell et al. (2009) recently proposed a multidimensional approach, based on six dimensions, extending the study of expertise also to groups, and how individuals may function within this groups. Below a table taken from their article, that resumes the 6 factors:

Table 3. Framework for the six dimensions of expertise

| ***Dimension*** | ***Content - Context - Process*** | ***Questions answered*** |
|---|---|---|
| *Subject matter* | *Content* | *What, (How)* |
| *Situational context* | *Context* | *When, Where, (Why)* |
| *Interface tools* | *Process* | *How* |
| *Expert identification* | *Content* | *Who, (When)* |
| *Communication* | *Process* | *What, How* |
| *Information flow paths* | *Context* | *Which, When, (How)* |

“Subject matter” is the classic form of expertise identified in domain knowledge context. “Situational context” is a general recognition of situational demands, as in situation awareness and situated cognition literature. “Interface tools” takes into account the expert’s skill to manipulate complex technological systems. “Expert identification” refers to the ‘know who’ social networking ability. “Communication skill” integrates communication, leadership and persuasion traditions. “Information flow path” expertise refers to the use of complex information and communication technologies to support physically and temporally distributed teams. Using these dimensions, the authors argue that a multidisciplinary viewpoint should be adopted, examining expertise shared across members of a group operating in a complex task environment. Their idea is that not only experts have to show the characteristics specific to a certain context, detailed in the table provided in the previous paragraph, but, when part of a team, must also show capabilities, such as effective communication skills, understanding of others limits and strengths, that can be assessed using a cross context approach.

It is possible contextualise the research described in this thesis using Garrett, Caldwell et al. approach. The participants to this research were Marine Pilots. Marine pilots are ship’s captains that are specifically trained and certified to manoeuvre vessels within critical coastal and port waters. This is the subject and the situational context of expertise specifically addressed in this study (subject matter). Pilots embark a ship outside port waters and then work with the bridge team, port services and facilities (tugs in assistance, Vessel Traffic Services (VTS) present in the area, linesmen etc..) to navigate and berth the ship (situational context). To be able to efficiently operate in this context, pilots need to be proficient in operating navigation and communication equipment (interface tools). Pilots need to effectively interact with crews. Even though on board of ships the hierarchical structure where the pilot is integrated is traditionally well defined (according to the prototypical pyramidal structure with the Captain on top) (Expert identification), multicultural and multi ethnical environments may pose challenges in how well messages and information may be received and addressed (communication skills). As anticipated, pilots, while working on board of the piloted vessel, need to interact also with other stakeholders, putting in place shared practises and procedures (Information flow path). In this study, some of these aspects, here just introduced, will be better described in chapter 3.

#### 2.1.2. Expertise and Memory

Memory, even if not directly investigated in this research, holds an important place in the model presented in Figure 2. Memory provides and holds the knowledge necessary to understand the specific shiphandling context and tasks. From the introduction of early experimental approaches to the human memory study (Ebbinghaus, 1967), it was clear that the most important factors influencing recall and retention were individuals' relevant experience, knowledge, and interests (Feltovich et al., 2006). For this reason, a laboratory based approach was initially designed to eliminate or, at least, minimize the effects of relevant experience, through the use of unfamiliar material. Without those effects, Ebbinghaus (1967) was able to study the basic laws of memory. Since then, simple tasks have been used by researchers to derive general laws and capacities for memory. This is how the traditional model of human memory was achieved. According to this model, memory is composed by a Working Memory (WM), a Short Term Memory (STM) and a Long Term Memory (LTM) (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965). The theory of Expertise (Newell & Simon, 1976) contributed to the extension of this model with additional mechanisms, able to explain observations about experts' greatly expanded working memory (F. S. Gobet, Herbert, 1996; Staszewski & Simon, 1995).

The study of memory processes in expert's domain of expertise, implied a critical shift from previous laboratory approaches. The information used and retained happened to be necessarily related to tasks, making very difficult to assess how many independent pieces of information were stored or maintained in attention and working memory. It was observing this capability (to relate pieces of information together), that suggested to Chase and Simon their explanation for experts' memory superiority (Herbert A. Simon & Chase, 1973). It became apparent how, through experience and practice, people learned to cognitively organize initially independent information, coming from their working environment, into larger units or "chunks". This way of integrating knowledge representations, made inapplicable or substantially attenuated the previously identified limits.

Studies carried out on chess experts (De Groot, 1978; Herbert A. Simon & Chase, 1973), showed how they were able to rapidly reorganize chess positions taken from a real game in already known patterns, while novices, not owning in their memory enough game configurations, were dealing with the board in a piece-by-piece manner. Experts were found to follow the same less effective novices' strategies, when confronted with pieces randomly placed on the board. Similar results were obtained in Bridge (Engle & Bukstel, 1978), GO, a traditional Chinese strategy board game (Reitman, 1976) and electronics (Egan, 1979). Besides, comparing these studies, it was also possible to confirm how in experts' memory, specificity plays an important role. Even if Go and Gomoko are played on the same board and use the same pieces, GO players showed quite poor performance in remembering Gomoko displays, and vice versa (Eisenstadt & Kareev, 1975). Even if novices and experts are constrained by the same short-term (or working) memory limitations (Cowan, Chen, & Rouder, 2004), they can rely on different strategies, with experts becoming capable to significantly increase their chunk size.

Differently than novices, experts do not simply rely on their transient short-term memory (Neil Charness, 1976), they are also capable to encode information in a long-term working memory LTWM (K. Anders Ericsson & Kintsch, 1995). This considerations apply also to experts in more physically dynamic fields, such as soccer (Postal, 2004), hockey (Weber & Brewer, 2003) and golf (Dijkstra, MacMahon, & Misirlisoy, 2008). Experts are able to anticipate potential future need, performing a skilled encoding in LTM of the information encountered and recognised as relevant. Such information is then automatically activated when the subsequent relevant contexts are encountered. A large body of relevant information becomes then available to experts, without any effort to actively maintain it in the limited STM (M. T. H. E. G. Chi, Robert (Ed); Farr, M. J. (Ed) 1988). Sun, Zimmer et al. (2011) showed how, for aviation pilots in a simulated environment, both spatial measures of WM capacity and LTWM skills were important predictors of situation awareness performance (SA) and their importance varied as a function of pilot expertise. Spatial WM capacity was most predictive of SA performance for novices, whereas spatial LTWM skill based was most predictive for experts. In addition Meade, Nokes et al (2009) studied how, non experts (non-pilots and novices) were relatively disrupted by mutual collaboration in a recall task, while experts showed to benefit from collaboration.

Even though not directly investigated in the context of this thesis, memory played a very important indirect role since:

- LTM and LTWM influence the ability of a marine pilot to plan. This ability was specifically explored in paper II (see sections 4.1.2 and 6.2).
- STM and LTM interact and support the pilot to build and maintain situational awareness. This aspect was investigated in paper IV (see sections 4.1.4 and 6.4).

Another fundamental mechanism achieved and exploited by experts to obtain their outstanding performance, compared to novices is Automaticity. Let see more in details how expertise and automaticity are connected.

### 2.1.3. Expertise and Automaticity

Shiphandling translates also into specific actions and into the execution of very specific tasks. Some examples could be provided by the effective use of equipment, such as radios and radars. Some practical tasks could include also interactions such as the clear explanation of manoeuvring intentions to the crew, use of standardised orders and language. All these tasks represent automated skills that pilots need to own to do their job. Proficient shiphandling might be considered as the flawless and skilful application of shiphandlers expertise into a very practical working environment - that is a vessel bridge with its crew and equipment as well as the surrounding shipping context.

As summarized by Anders, Ericsson et al. (2006), in the traditional theory of expertise (P.M. Fitts & Posner, 1967), skill acquisition is progressive, starting with the acquisition of a cognitive task representation and a certain number of possible reactions in typical situations. With subsequent practice, actions become smoother and more efficient, till the stage when finally, performance is achieved with a minimal effort, running essentially automatically, without active cognitive control (Posner & Snyder, 2004). Rasmussen differentiates between three levels of human performance: skill-based (a sensory motor performance taking place without conscious control), rule-based (internally stored rule, which requires limited control while achieving a familiar goal), knowledge based performance (higher level which requires conceptual control since the performance is novel and or the context is unknown) (Rasmussen, 1983). Thanks to the reduction in the cognitive demand due to automaticity, more processes can then run in parallel (Shiffrin & Schneider, 1977). For example expert typists can type and recite rhymes at the same time (Shaffer, 1975). Skilled abacus operators could answer routine questions without loss of accuracy or speed in working with the abacus (Hatano, Miyake, & Binks, 1977). Students, after practice, could read unfamiliar text while simultaneously copying words read by an experimenter, without decrement in reading speed or comprehension (Spelke, Hirst, & Neisser, 1976). Each automated process is also individually more resilient to disruption by reduced cognitive capacity (Schneider & Fisk, 1982). In a study on golfers, novices performed worse under instructions to putt as quickly as possible, relative to instructions that did not limit execution time, while the opposite was true for experts. Novices needed time to attend and control performance, while performances of experts appeared to be adversely affected by unlimited execution time. Time availability seemed to allow the counterproductive opportunity to explicitly attend to and monitor automated execution processes (Sian L. Beilock, Bertenthal, McCoy, & Carr, 2004).

There are some concerns about adopting this general model to explain expertise (K.A. Ericsson & Smith, 1991; Holyoak, 1991; T.A. Salthouse, 1991). There is evidence, in fact, that in experts, complex cognitive representations continue to mediate performance and learning (K.A. Ericsson, 1996, 2003). Nevertheless automaticity is important, since only when basic cognitive components such as decoding, encoding of input, are automated, they can allow higher level skills such as reasoning, comprehension, inference, monitoring, and integration to be proficient (Logan, 1985).

Longitudinal studies have shown the effects of this progression. In a group of children, researchers found that if basic reading skills did not become automated, comprehension skills could not substantially develop (Lesgold & Resnick, 1982). Furthermore, speed increases in word skills, predicted later comprehension increases and not the opposite. Interactions between automaticity and the capability to recall available knowledge, have already been described in the previous paragraph, associated with overload or inefficiency in the use of working (or short-term) memory (Jeffries, Turner, Polson, & Atwood, 1981; Johnson et al., 1981).

In this thesis, the effects of automaticity were indirectly witnessed when considering pilot performance. Paper II (see sections 4.1.2 and 6.2) proposed a comparison between plans and execution. Participants' execution of their manoeuvres in the simulator was a demonstration of automaticity in action. Pilots, while monitoring the development of the ship motion, were counteracting undesired effects giving rudder and helm orders (as well as instructions) to crew and tugs. To be clearly understood by bridge teams of any nationality, these orders and instructions have to be in standard English, meaning that sentences and words have to adhere to a standard vocabulary (IMO, 2001). The translation of intentions into orders (on the shiphandler side) as well as their execution (on the crew and tugs side) has to be prompt and accurate, implying the existence of acquired and well-practised skills. In this study, it was documented how gaze scanning strategies require similar levels of automaticity. As described in paper IV (see sections 4.1.4 and 6.4) it was demonstrated how shiphandlers, depending on the manoeuvring context, adopt different scanning behaviours organised in cyclical gaze sequences. From the literature, it is known that the acquisition of automated skills requires and induce deep changes at a neuronal level. Different parts of the brain are involved during acquisition of a novel skill compared to those involved in the execution of a familiar task. Using fMRI it was possible to highlight how increased automaticity correlated with decreased cerebral activity, in the same anatomical loci that showed higher activity during acquisition (Jansma, Ramsey, Slagter, & Kahn, 2001). Studies in perceptual categorization, showed that early performance is dominated by higher cerebral activity in sub cortical paths. These paths are characterized by greater neural plasticity thanks to dopamine-mediated learning signals from the substantia nigra (a basal ganglia structure located in the midbrain), while when automaticity is achieved brain activity is characterized by faster, cortical-cortical projections (Ashby, Ennis, & Spiering, 2007).

### 2.1.4. Expertise and Attention

One of the principal duties that pilots are expected to fulfil once on board, is to monitor vessel progress while sailing in critical waters. This requires pilots to constantly shift their attention among the several tasks involved. Rosenbaum, Augustyn et al. (2006) demonstrated how the differential involvement of attention in skilled performance was previously considered in the three stage model of Fitt and Posner. The model described the acquisition of skills from an intellectual cognitive stage through an associative procedural stage, to an automatic independent stage. The model accounted that the progress in skill acquisition was accompanied by a decreased reliance on conscious cognitive control and focused attention (P.M. Fitts & Posner, 1967). The same concept was supported in the later Shiffrin and Schneider model where the authors made a distinction between controlled and automatic processes (Shiffrin & Schneider, 1977). In their view, controlled processes required attention, whereas automatic processes did not. The critical role of attention during early perceptual-motor skill acquisition in a sequence-learning task was also shown by Nissen and Bullemer. Participants in a dual-task condition had inferior results learning the sequence, compared to the single-task condition (Nissen & Bullemer, 1987). Being expert or novice in the performed activity makes the difference. Skilled pianists could accurately read musical scores, while repeating words presented through headphones (Allport, Antonis, & Reynolds, 1972). In sport, novice hockey players showed a decrease in their basic hockey skills when simultaneously involved in a visual shape-monitoring task while expert hockey players were mainly unimpaired (Leavitt, 1979). Expert golfers, did not decrease their performance while simultaneously monitoring a string of verbally presented words, while novices putt less accurately during the secondary task and recognized fewer of the monitored words (Sian L Beilock, Wierenga, & Carr, 2002).

Experts can therefore avoid focusing on the immediate performance and direct their attention more efficiently towards other elements and forthcoming consequences. Even for golf novices, focusing attention on more distant details of their performance (paying attention on swinging their club instead of swinging their arms) was associated with more accurate shots (Wulf, Lauterbach, & Toole, 1999). Novice tennis players had more accurate forehand shots when they focused on the trajectory of the ball instead of on its contact with the racket (Maddox, Wulf, & Wright, 1999). A number of studies support the conclusion that skill acquisition seems to be facilitated by focusing on the consequences rather than on the actions themselves (Wulf & Prinz, 2001). Attending with too much emphasis to the single elements of a skill may result in an over-regulation of muscular degrees of freedom, limiting the ability to effectively implement motor plans (Riley, Stoffregen, Grocki, & Turvey, 1999). Paradoxically, too much attention can interfere with expert performance and “Choking” (failure due to poor skill execution when the individual is close to winning) could be a classic example how someone’s mind can interfere with performance, as shown in golfers and footballers (Sian L Beilock, Carr, MacMahon, & Starkes, 2002). Mechanisms for self-regulation of arousal level, thoughts, performance expectancy, and attentional focus take place before and during a sporting event. Singer (2002), in his review on self-paced sports such as archery and shooting, offers an insight how psychophysiological indices such as visual gaze, EEG activity, and heart rate measures, collected pre and during the performance routines, could be used to characterize an expert performer profile. Not only too much attention but also too many detailed instructions can interfere with experts’ capability to deal with tasks. In their review, Kalyuga, Rikers et al. (2012), detailed how less experienced students may in some situations outperform seasoned medical practitioners on recall of specific cases. More experienced technical trainees or students may learn less than expected from instructions that are very effective for novices.

### 2.1.5. Expertise and Information Selectivity

While conducting a vessel, pilots are constantly exposed to a continuous flow of information such as data provided by bridge equipment, crew communications, radio calls, different ships moving in the area as observed through the bridge windows, etc.. All these sources of information may or may not be relevant to the specific task at hand. These elements, though, become part of pilots’ awareness and prompt a constant comparison with pilots’ owned repertoires. Selectivity addresses the issue of how experts are able to efficiently access their structures of knowledge and the relevant information taken from the context (Feltovich et al., 2006). Both in the choice of relevant events and memories, a degree of appropriate abstraction is required, in order to identify features as familiar. Experts exploit, to classify problems, those abstracted and not necessarily apparent features (M. T. H. Chi, Feltovich, & Glaser, 1981). They develop hierarchical organizations, typical of experienced memory (Feltovich, Johnson, Moller, & Swanson, 1984; Patil, Szolovits, & Schwartz, 1981). Selectivity is based on the attribution of differential importance to the features extracted from events or to internal cognitive processes themselves. Selectivity was developed as task adaptation, due to human limitations in cognitive capacity. While experiencing different types of situations, humans selectively search for abstract invariances and discriminating cues, later considered integral to a specific task. Expertise, then, involves learning which information is most useful and which is superfluous (M. T. H. Chi, Feltovich, et al., 1981; Hinsley, Hayes, & Simon, 1977; Patel & Groen, 1991; Spilich et al., 1979). In less complex environments (or tasks), the important invariance can be well defined so that the link between selectivity and performance can be easier explicated. An example could be provided by skilled typists: developing integrated representations of letters and key presses, they strongly facilitate the transit from perception to physical response (Rieger, 2004). At a lower perceptual level, studies about "Change Blindness" showed how results varied if the changed detail was relevant to field of expertise of involved subjects. "Change Blindness" is the incapability of an individual to detect a change between almost identical visual stimuli, scene, pictures, or, as in the case of the mentioned study, physics diagrams (Feil & Mestre, 2010). More broadly, Expertise is also involved in knowledge inversion. In knowledge inversion, experts have to move from a general concept to contextual problem details, finding regularities, integrating exceptions and accepting natural variations, judging if encountered features are still pertinent to the relevant concept. Medicine could offer a good example, since medical students have to build up their "disease" knowledge, the underlying pathophysiology, variations, and classic manifestations and then they are faced with a patient. According to this "case oriented learning", in which medical students are given early exposure to representative clinical situations, learners are forced to develop mental representations and a LTWM that support medical reasoning under real-time, representative constraints.

### 2.1.6. Expertise and Experience

In this doctoral study it was assumed that participants were similar in term of experience. Participants were selected from the same Pilot Company and it was used a structured interview to evaluate and document participants' previous relevant experience. With reference to this research assumption, this paragraph expands on the relationship between expertise and experience, as documented in the literature. As Ericsson (2006b) summarizes, individuals are able to learn in less than 50 hours, what they need to know, to obtain a functional level of performance in a general activity. After this time, tasks are sufficiently automated, so that not much attention is required (Anderson, 1982, 1987; P.M. Fitts & Posner, 1967; Shiffrin & Schneider, 1977). According to some authors, expertise can be a direct consequence of lengthy experience, so that individuals with over ten years of full-time engagement could be considered experts in a domain. Expertise is mediated by instruction, training, and experience so criteria for identifying experts could be based on social reputation, education, accumulated accessible knowledge, and length of experience (M. T. H. E. G. Chi, Robert (Ed); Farr, M. J. (Ed) 1988; R.R. Hoffman, 1992). Over ten years of practise are necessary in several domains before an individual could gain international level recognitions (Herbert A. Simon & Chase, 1973). The "ten years rule", though, has to be considered more a necessary than a sufficient condition to expertise and it seems to have found confirmation in a wide range of domains: chess (Herbert A. Simon & Chase, 1973) music (Hayes, 1989; Sosniak, 1985), mathematics (Gustin & Bloom, 1985), tennis (Monsaas, 1985), swimming (Kalinowski & Bloom, 1985), long-distance running (Wallingford, 1975), science and novel writing

(Raskin, 1925). It is also true, that once an acceptable level of performance has been reached, it can be kept stable with minimal effort for years and even decades. This is why, the length of experience beyond the first two years in many studies shows a weak correlation with job performance (McDaniel, Schmidt, & Hunter, 1988). Clinical psychologists' professional experience, for example, did not strongly correlate to their treating success (Dawes, 1996). Software design experience did not show consistently superior proficiency (Rosson, 1985; Sabine Sonnentag, 1998), as well as wine experts' did only slightly better than regular wine drinkers (Gawel, 1997; Valentin, Pichon, de Boishebert, & Abdi, 2000). Financial experts' decisions and forecasts, did not show a reliable superiority over novices' ones (Camerer & Johnson, 1997; Shanteau & Stewart, 1992). Similar results were found in computer science (Doane, Pellegrino, & Klatzky, 1990), and physics (Reif & Allen, 1992). A study on expert and novice nurses, investigated their capability to assess patients, using a thinking aloud protocol. Although experts possessed and accessed more basic and subordinate concepts than their novice peers, the inclusivity and complexity of their concepts appeared to be the same (Greenwood & King, 1995). In addition, watching an activity will not be as effective as being directly involved. Large differences in the ability to anticipate events in soccer, were found between players and avid spectators (M. Williams & Davids, 1995). There are even paradoxical examples: in physicians' diagnosis of heart sounds and x-rays (K.A. Ericsson, 2004) and auditor evaluations (Bédard, Chi, Graham, & Shanteau, 1993), performance decreased systematically with the length of professional experience, after the end of formal training. It seems that once an acceptable level is attained, there are hardly any benefits from the common kind of additional experience. Additional experience contributes making performance less effortful and demanding. To improve performance, though, specific activities aimed to improve dedicated skill aspects are necessary. Carrying out these activities in a supervised and protected environment, the trainee experiences the opportunity to evaluate alternatives, as well as to perform methodical repetitions with informative feedback. This type of activity falls under the name of Deliberate Practice.

### 2.1.7. Expertise and Deliberate Practise

Pilot companies – like all high risk organisations, every year invest significant time and resources in training. Training, seen as the engagement in specific and relevant practices to improve and reach certain levels of performance, is considered one of the fundamental paths to safety. The ideas that engaging in practice necessarily leads to maximal performance as well as the conception that innately talented individuals can easily and rapidly achieve an exceptional level of performance were disproved by biographical evidence (K. Anders Ericsson et al., 1993). Different types of experiences have qualitatively and quantitatively different effects on the continued acquisition and maintenance of an individual's performance (K.A. Ericsson, 1996; K. Anders Ericsson, 2002; K. Anders Ericsson et al., 1993). This became apparent since early studies on Morse Code operators (Bryan & Harter, 1897, 1899) where plateaus in skill acquisition, could be overcome only with extended specific efforts in dedicated training (Keller, 1958). As seen in the previous paragraph, at least ten years of experience are generally required to achieve an expert level of performance in a field. Ericsson (2006b) underlines how world-class levels seem to be reachable in less than ten years only in those activities lacking a history of organized international competitions. But how does, this long path eventually conducting to excellence, start and develop? From retrospective interviews of experts in many domains, a common finding was that elite performers were introduced to their field of excellence at a young age (Bloom, 1985). Generally a good support was offered by their parents, investing a considerable amount of time and resources, to help them finding good coaching and allowing them to benefit of regular practice at the best training centres (Bloom & Sosniak, 1985). Yet, the best training environments may not be sufficient to produce the very best performers. Engagement in "Deliberate Practise", individual commitment to activities specifically dedicated to improve certain aspects of performance, it is what seems to differentiate experts from other learners.

With deliberate practice (K.A. Ericsson, 1996; K. Anders Ericsson, 2002, 2004; K. Anders Ericsson et al., 1993) the expert performance is acquired gradually, sequentially and progressively under the supervision of a teacher or a coach. Tasks initially outside current learner capabilities, are acquired with hours of practice concentrating on critical aspects and through repetitions and feedback. Study methods, consistent with deliberate practice, were able to predict achievement in both undergraduate college students (E. A. Plant, Ericsson, Hill, & Asberg, 2005) as well as in students in medical school (Moulaert, Verwijnen, Rikers, & Scherpbier, 2004). In a study at the music academy in Berlin, expert violinists' activities were recorded on a diary (K. Anders Ericsson et al., 1993). Even though all violinists spent about the same amount of time (over 50 hours) per week on training, the best among them dedicated more of this time to deliberate practise. Even "talented" children spend more time in deliberate practice each week (Sloboda, 1996). In sports, where general performance has continually improved in time (K.A. Ericsson, 1990; Schulz & Curnow, 1988) and where individual peak performance is nearly always attained many years after initial exposure (Schulz & Curnow, 1988), a consistent relationship between attained performance and amount of deliberate practice is present (Helsen, Starkes, & Hodges, 1998; Hodges & Starkes, 1996; Starkes, Deakin, Allard, & Hodges, 1996). Similarly in chess, the amount of solitary study is the best predictor of chess skill, with only a very small benefit from games played in tournaments (N. Charness, Krampe, & Mayr, 1996; N. Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005). Similar findings were found in darts (Duffy, Baluch, & Ericsson, 2004).

Moreover, successful handling of emergencies by airline pilots was correlated with practise of the same emergencies in the simulator (McKinney & Davis, 2003). Although the detailed nature of deliberate practice will differ across domains and as a function of attained skill, there appear to be limits on the daily duration of deliberate practice, and this limit seems to be generalized. When individuals start practicing, the amount of practice is an hour or less per day (Bloom, 1985). Even expert performers from many domains engage in practice without rest for only around an hour (K. Anders Ericsson et al., 1993). In elite musicians (K. Anders Ericsson, 2002) and athletes (K Anders Ericsson, 2001; K.A. Ericsson, Starkes, & Ericsson, 2003) the limiting factor seems to be the inability to sustain the necessary level of concentration. In many diverse domains the amount of practice never consistently exceeds five hours per day (K. Anders Ericsson et al., 1993; Krampe & Ericsson, 1996), as shown in writing (Plimpton, 1977). So even if in real life, practice durations can range from 1 to 8 hours per day, no considerable benefit has been shown from exceeding 4 hours per day and only reduced benefits from practice exceeding 2 hours (Welford, 1968; Woodworth & Schlosberg, 1954).

Deliberate practise helps expert performers to continue their development despite automaticity. By continuously stepping up to more demanding tasks, they stretch their learning further. Their acquisitions will reflect the demands of the particular activity they engaged themselves in, and thus will differ from one domain of expertise to another, even though the overall structure of these mechanisms may reflect general principles. It is also important to remember that deliberate practise is a privileged path not only to gain but also to retain expertise, especially on practical tasks and abilities. A simple example could be provided by a study, carried out on more than six hundred nurses. The study tested their capability to retain over time cardiopulmonary resuscitation techniques, following the initial training. Students who had deliberate practice on their compression, ventilation, and single rescuer skills on voice advisory manikin, had better performance than a control group with no practice beyond the initial training (Oermann et al., 2011). In the recent years, a critical achievement in the way deliberate practise can be pursued, has been reached through the use of simulators. As an example, McGaghie, Issenberg et al. (2009) offer a comprehensive review of the most important features developed by Simulator Based Medical Education in the recent years. In their work, which looks back into medical history more than 40 years, they were able to list 12 different areas where medical training is currently benefiting from the use of simulator (such as team training, skill acquisition and maintenance, etc..).

In the maritime industry, in addition to computer simulators, deliberate practise in shiphandling is also achieved through the use of “manned models”. These models are ships in miniature (length overall generally less than 10 meters) with hull shapes, propulsion and steering appropriately scaled to reproduce the inertia and the manoeuvring behaviour of bigger vessels. Shiphandlers would be sitting into these models and would practice the use of the hydrodynamic effects, repeating specific manoeuvres in dedicated facilities.

Even though not yet fully integrated in international certifications (STCW, 2011), requirements to demonstrate continuous professional development are gaining momentum at a National level. In Australia, for example, the efforts of professional associations (Australasian Marine Pilots Institute) in conjunction with State and Federal Regulators (Maritime Safety Queensland, Australian Maritime Safety Agency) are aiming to endorse dedicated training paths able to ensure that professionals in the sector maintain a suitable level of preparation over time (www.ampi.org.au/cpd). As will be better detailed in chapter 5, one of the aims of this research was to develop a methodology able to demonstrate how training standards could be defined and how to evaluate if those standards were met.

### 2.1.8. Expertise and Decision-Making

In this research, as depicted in Figure 2 in section 1.2., decision making is simply described as the mechanism through which pilots, after comparing their expectations with the experienced outcome, trigger their future actions. This paragraphs expands on the topic as presented in the literature, providing additional elements. Yates and Tschirhart (2006) underlined how it is sometimes a mistake to consider decisions as “good”, only when they lead to desirable outcomes (J. F. Yates, Veinott, & Patalano, 2003). An outcome bias happens (Vlek, 1984) when people, basing judgement on information that is only available after the decision is made, take outcomes into account, instead of the quality of the decisional process itself (Baron & Hershey, 1988). According to this view, decisions should be judged based on whether or not the decider was self-contradictory in the process (J. F. Yates, 1990). In a perhaps extreme version of this perspective, Edwards contended that the sole criterion for decision quality should be whether the process, used to arrive at a decision, follows the maximization of expected utility (Vlek, 1984). Studies on decision making were initially conducted mainly in laboratories and uncertainty was not explicitly acknowledged (Payne, Bettman, & Johnson, 1988). As summarized by Klein, Shafer et al. (2006), two basic paradigms of research on decision-making, adopted this approach: the formal-empiricist paradigm (also known as classic decision making CDM) and the rationalist paradigm (M. S. Cohen, 1993). The formal-empiricist paradigm is a normative and prescriptive model of rational behaviour, where the decision maker chooses among concurrently available alternatives. There is an input-output orientation, a comprehensive information search and a formal development of an abstract, context-free model, suitable for quantitative testing (R Lipshitz, 2001). Although the focus was on behavioural testing, the effort was mainly on imposing constraints of mathematical consistency on a subject's judgment and not on understanding cognitive processes.

The rationalist paradigm tried to recover from this weakness, though still retaining the essential characteristics of a normative and prescriptive model. In the rationalist paradigm errors were biases due to unaided decision making. Discrepancies in performance were decision maker fault, not flaws in the model (M. S. Cohen, 1993). The rationalist paradigm, compared to the formal-empiricist one, was more focused on cognitive aspects of the decision, exploring intellectual limitations, revealing psychological processes, and introducing the use of intuition (Kahneman & Tversky, 1982). On one hand, the formal-empiricist paradigm combined normative and descriptive functions in the same formal models, on the other hand the rationalist paradigm separated the two functions, cognitively explaining the studied behaviour and formally evaluating it. Nevertheless, both the above paradigms referred to classical decision mathematical models, analysing decision making from the perspective of game theory (Von Neumann & Morgenstern, 2007) and using statistical models to demonstrate decision biases (Slovic, Fischhoff, & Lichtenstein, 1977). As highlighted by Salas, Rosen et al. (2010), it is a general approach detectable across many authors, conceptualizing human information processing in terms of two distinct systems (Chaiken & Trope, 1999; Evans, 2008, 2009; Moskowitz, Skurnik, & Galinsky, 1999). To cite some examples of dual process theories: automatic and controlled (Shiffrin & Schneider, 1977), experiential and rational (Epstein, 1994), holistic and analytic (Nisbett, Peng, Choi, & Norenzayan, 2001), reflexive and reflective or X and C systems (M. D. Lieberman, Jarcho, & Satpute, 2004), associative and rule-based (Sloman, 1996) conscious and unconscious (Dijksterhuis & Nordgren, 2006), intuitive and analytic (Hammond, 1996).

Although there are important distinctions, many are the similarities among these models, they all describe a first system that is fast, holistic, and does not require conscious cognitive effort (the intuitive system) and a second system that is slower, analytic, and cognitively effortful (the conscious deliberative system). Stanovich (1999; 2000) and others (Evans, 2008) referred to System 1 as the intuitive system and System 2 as the conscious deliberative system. Unlike what happens in laboratories, in real life the results of almost every action depend, at least partly, on events outside decider’s control, awareness and anticipation. A typical example, mentioned by Kitson (1999) and documented by Cranley, Doran et al. (2009), is the medical industry, where one of the most important skill is the ability to recognize and handle clinical uncertainty. Starcke and Brand (2012) state that if uncertainty exists and a rational mathematical calculation of choices is not possible, then both the previously described systems are needed. A temporary automatic response coming from System 1 may anticipate the slower reacting deliberate System 2, allowing adjustments while making inferences (Gilbert, 1999). In the presence of some degree of uncertainty or a conflict between emotional intuitive and deliberate strategic decisions, both systems may act together (Greene & Haidt, 2002; Haidt, 2007; Yamagishi et al., 2009). Also feelings may take part to the Decision Making process.

Pham (2004) argues that feelings can serve as proxies for values, they can be used as information source for alternatives, they also prime thoughts, triggering contents to consciousness. Neuropsychological decision-making research focuses on emotions associated with alternatives and subsequently associated with decisions (Bechara, 2004). Damasio, Everitt et al. (1996), with the somatic marker hypothesis, postulate the connection between feedback processing and decision making. Decisions in uncertain situations are guided by somatic markers. The somatic markers come from experiences of a reward or a punishment, when emotional responses associated with a decision were linked to certain somatic states. The same somatic states are then re-experienced during a current decision and will have effect on the actual available alternatives, indicating which of them should have resources allocated in the working memory. These markers act as starting or warning signals, guiding decisions. Dunn et al. (2006) provide a critical evaluation of the somatic marker hypothesis. Decisions that are made under high uncertainty, in which are present no other clues besides future feedback, are particularly sensitive to reward and punishment (e.g., gambling). Decision situations can be described on a continuum from high to low degrees of uncertainty. The placement on this continuum, likely triggers specific decision-making mechanisms, such as application of familiar strategies, adjustment from automated responses, feedback processing and reliance on reward and punishment (Brand, Labudda, & Markowitsch, 2006). In their review, Starcke and Brand (2012) investigated how stress interacts with and effects decision making, describing also its connections on a neuronal level. They described also how stress may interfere with executive functions, such as working memory, exaggerating reliance on lower level automatic responses and decreasing control of cognitive processes.

However beneficial the traditional decision-making research carried out in laboratories was for building understanding of aspects of human cognition and choice, it did not provide enough understanding of how professional judgment and decisions were taken in the field (G Klein, Orasanu, Calderwood, & Zsambok, 1993). As explained by Klein, Shafer et al. (2006), research had to move into the real world. The focus of Naturalistic Decision Making (NDM) research centred on expert practitioners (Raanan Lipshitz, Klein, Orasanu, & Salas, 2001), while they were attempting to make decisions under difficult circumstances (G Klein et al., 1993). Key contextual factors were now taken into account, such as: ill-structured problems, uncertain and dynamic environments, shifting and ill-defined or competing goals, time constrain and stress, high stakes with multiple players and organizational goals and norms involved (Zsambok & Klein, 2014). Early NDM research discovered that experts expend a considerable effort on situation assessment, then evaluate single options through mental simulation, and then arrive at a satisfactory answer or action. The previously mentioned laboratory models of decision making, were not of much help in understanding these new findings. Specifically, in the utility theory, the decision maker lays out all of the alternative decision paths and iteratively evaluates each for costs and benefits to reach a good judgment. The successful professional judgment being observed in the field was radically different from these prescriptive processes.

The decision maker was the distinguishing focus of NDM (Raanan Lipshitz et al., 2001). Expertise became a core of NDM research and being investigated from a number of theoretical perspectives, such as in the Judgement and Decision Making (JDM) literature. JDM studied heuristics and biases in an attempt to demonstrate biases among expert populations (Kahneman, Slovic, & Tversky, 1982). Experts do not perform well when the tasks depart from their familiar tasks, while when experts perform in their domain or in their natural context, biases are alleviated and they show good judgments (Bornstein et al., 1999; M. S. Cohen, 1993; Keren, 1987; Shanteau, 1989; J. F. Smith & Kida, 1991).

A substantial literature devoted to capture expertise across many domains developed thanks to the advent of computer applications and expert systems (Boose, 1986; Bramer, Bramer, & Bramer, 1985; Coombs, Dawes, & Tversky, 1970; Robert R Hoffman, Shadbolt, Burton, & Klein, 1995; Waterman, 1986; Weiss & Kulikowski, 1984). The attempt was to try to capture an objective expertise model, as if there was one ideal for a given domain. When NDM researchers studied experts, they meant individuals who had achieved exceptional skill in one particular domain, and the NDM research focused on understanding the process of developing and applying that expertise in a context. Researchers have defined a number of variables of expertise that are important to NDM researchers, as already encountered in previous paragraphs (Phillips, Klein, & Sieck, 2004). An integrated table (G Klein et al., 2006; Salas et al., 2010) provided below, summarizes the main mechanisms involved in expert decision making as reported in the literature by different authors:

Table 4. Main mechanisms involved in expert decision making

| Mechanism | Description | Key points | Citation |
|---|---|---|---|
| *Large and well-developed knowledge base* | *Expertise-based intuition uses knowledge. Conceptual and procedural knowledge are aspects of this knowledge base.* | *Experts organize knowledge in a conceptual way.*<br>*Experts organize knowledge with more interconnections between concepts.*<br>*Knowledge is organized via semantic networks, theories, and schemas.* | (Bordage & Zacks, 1984)<br>(M. T. H. Chi, Ohlsson, & Holyoak, 2005)<br>(Feltovich et al., 1984)<br>(Markman, 1999) |
| *Tacit knowledge* | *Tacit knowledge is the operational knowledge inaccessible to consciousness.* | *Experts know more facts and details and have more tacit knowledge than novices do.* | (BW Crandall, Kyne, Militello, & Klein, 1992) |
| *Perceptual skills* | *Ability to make fine discriminations among different stimuli coming from the environment.* | *Experts see more in a situation than a novice, by noticing cues a novice does not recognize.* | (Gary A Klein & Hoffman, 1993) |
| *Sense of typicality* | *Ability to perceive the situation as familiar* | *Experts recognize when things are not going as expected, that is, when there is an anomaly or something is missing.* | (K. Ericsson & Simon, 1993) |
| *Pattern recognition* | *Expertise-based intuition uses a collection of complex patterns in a person's domain to perceive larger and more meaningful patterns in the environment more rapidly.* | *Experts view cues as chunks or patterns.*<br>*Experts' use of pattern recognition allows them to assess the environment more rapidly than novices.*<br>*Affords the ability to use pattern matching effectively.* | (S. F. Biggs & Wild, 1985)<br>(Eggleton, 1982)<br>(F. S. Gobet, Herbert, 1996)<br>(Neisser, 1976)<br>(Herbert A. Simon & Chase, 1973) |
| *Sense making* | *The effort exerted to understand events in order to create order and make sense of what has occurred, what is occurring, and what will occur.* | *Experts engage in problem detection. identification, anticipatory thinking. forming of explanations, identifying explanations, discovering inadequacies in initial explanations. and projecting the future.* | (Gary A Klein, 1993)<br>(Gary Klein, Phillips, Rall, & Peluso, 2007)<br>(Weick, 1993)<br>(Weick, 1995) |
| *Situation assessment and problem representation* | *Expertise-based intuition utilizes situation assessment and problem representation, which includes maintaining an understanding of the entire picture.* | *Quick judgments can be made of the situation (e.g., atypical or familiar).*<br>*Identification and clarification of the state of a problem.* | (Mica R Endsley, 1995)<br>(Randel, Pugh, & Reed, 1996)<br>(Mosier, 1991)<br>(Flin, Slaven, & Stewart, 1996) |
| | | *Experts spend more time than novices understanding the dynamics of the situation* | (Kobus, Proctor, Bank, & Holste, 2000) |

| Mechanism | Description | Key points | Citation |
|---|---|---|---|
| | | *Novices spend more time deliberating over the course of action* | |
| *Automaticity* | *The process by which an individual can accomplish a task without using all cognitive resources.* | *Accomplishing a task is not affected by or affects a concurrent task.* | (Schneider & Shiffrin, 1977) |
| | | *Contributes to an expert's ability to understand the larger meaning of a set of events.* | |
| *Managing uncertainty* | *Capability to face uncertainty* | *Experts use a range of strategies for managing uncertainty in the field* | (Raanan Lipshitz & Strauss, 1997; Schmitt & Klein, 1998) |
| *Mental models* | *Internal representations of how things work in expert's domain of practice* | *These mental models allow them to learn and to understand situations more rapidly* | (Rouse & Morris, 1986)<br><br>(K. G. Ross, Battaglia, Phillips, Domeshek, & Lussier, 2003) |
| *Finding leverage points* | *Leverage points are opportunities for making critical changes at relatively low effort* | *Experts can find leverage points in a situation and capitalize on them to implement innovative strategies* | (G. Klein & Wolf, 1998)<br><br>(K. Ross et al., 2002) |
| *Mental simulation* | *Provides an evaluation of a course of action to a situation, specifically if the course of action '"fits" the situation.* | *Conscious and deliberate process.* | (Gary Klein, 2008)<br><br>(Rutherford & Wilson, 1989)<br><br>(GA Klein & Crandall, 1995) |
| | | *Affords the ability to engage in simulated implementation of the solution.* | |
| | | *During this process the decision maker evaluates the quality of the solution.* | |
| *Metacognition* | *Understanding one's own strengths and limitations* | *Experts are better self-monitors than novices* | (M. T. H. Chi, Feltovich, et al., 1981; Larkin, McDermott, Simon, & Simon, 1980) |

Today, NDM encompasses a number of models and theories about how expertise works. Among these models Klein (1993) described the Recognition-Primed Decision (RPD) developed from the observations of fire ground commanders (Gary A Klein, Calderwood, & Clinton-Cirocco, 1986). The RPD Model states that when it comes to high-stakes, time-pressured decisions, experts do not use "rational choice" or utility analysis, instead, they rely on their experience, recognizing the situation as typical, as a prototype. This prototype brings also what to expect from the situation (expectancies), suitable goals, typical Courses Of Action (COA), and relevant cues. Exploiting this prototype the expert doesn't need to go through elaborate analyses. This expert initial recognition can lead directly to action with no comparison of options. The expert already knows that the COA will work. Often in the field there is no time to seek the optimal solution, so all what is required from RPD is a solution that will work, according to a "satisfying" prospective (H A Simon, 1957). If a situation is unusual or uncertain enough, that the predefined COA needs evaluating, mental simulation will be required. Mental simulation is the process of consciously envisioning a sequence of events, allowing a decision maker to make accurate predictions about the consequences of a particular COA. The expert will seek a COA that will meet his goals and will fit the constraints of the situation. The production of COA is sequential, never comparing options against each other. The expert ability, developed through experience, depends on the skill at recognizing typical situations. The Critical Decision Method (CDM) ((Gary A Klein & Brezovic, 1986), p. 17) is a knowledge elicitation technique that was developed in tandem with and in order to study the RPD Model and was employed to study experts in different fields (Robert R Hoffman, Crandall, & Shadbolt, 1998).

Klein, Shafer et al. (2006) provide several examples: In the medical field, Crandall and Calderwood (1989), studying highly experienced neonatal intensive care unit (NICU) nurses, found that they relied heavily on the recognition of perceptual cues to identify sepsis early stages. In chess, Klein, Wolf et al. (1995) found that the first move considered was of significantly higher quality, as judged by a chess Grand Master, than would be expected from a random sample of available moves. In addition, highly skilled players generated high quality moves even under time pressure, differently from medium-skilled players that did significantly worse (Calderwood, Klein, & Crandall, 1988).

A military study on novices and experienced tank platoon commanders (Brezovic, Klein, & Thordsen, 1990), demonstrated how students, even though provided with the same cues of instructors, considered alterative hypothetical courses of action less than instructors. Students were also less likely to recall hypothetical actions or situation features in a decision point. Students were less recognitional, as expected, since they were inexperienced, and this confirmed the hypothesis that people deliberate when they lack the experience to do RPD. The difference between experts and novices was not in strategies but in knowledge. Other examples drawn from the military are the Military Decision Making Process (MDMP) (USArmy, 1997) and the U.S. Marine Corps Planning Process (MCPP) (USMarineCorps, 1998). The MDMP and MCPP are highly proceduralised and cumbersome to employ (Gary Klein, 1997; Raanan Lipshitz, 1993). A prescribed analytic process does not help military decision makers in the field. Expert Officers generally are able to satisfactorily assess a situation, even if not exactly the same as others previously encountered. When the typical aspects of a situation are recognized, a plausible COA usually comes to their mind. Experienced decision makers then assess that course of action by mentally war gaming it, rather than contrasting it to other options on a set of abstract evaluation dimensions as required by the MDMP or MCPP (Fallesen & Pounds, 2001; G. Klein, 1999; Gary Klein, 2004; Pascual & Henderson, 1997).

Schmitt and Klein (1999) developed the Recognitional Planning Model (RPM) where the commander can identify his preferred COA and the staff can work on detailing and improving it. The intent was to codify the existing and effective practices and give the military a set of procedures that reflect their best practices as these have evolved over decades (K. G. Ross et al., 2003) p. 5). In engineering, Klein and Brezovic (1986) conducted CDM interviews with professional system designers. Even if the decisions elicited were not as time-pressured as those of fire fighters, the highest frequency of decision-making strategy was recognitional. The designers tended to avoid formal decision making, being less interested in finding the best option possible. They preferred to identify the best readily available option and make it more effective. In the offshore Industry, Flin, Slaven et al. (1996) found that the most experienced managers had emergency response schemata in place, that they used to assess incidents and recognition-based rules that they used to manage those incidents.

In the context of shiphandling, decisions are made at different times and levels. One of the first and most important decisions that is made is related to the safety and the feasibility of a manoeuvre. Many are the elements that are considered in this decision: features of the geographical area, port infrastructures and facilities, type and dimensions of vessels, loading conditions, environmental conditions related to wind, current, tide, visibility, just to mention a few. Any responsible authority for a specific area, would normally establish safety limits beyond which operations would not be allowed. Even though a manoeuvre may start in compliance with those limits, changes in environmental conditions may very quickly put the vessel in a completely different scenario, at a time when decisions could not be withdrawn (vessel already within port limits). One of the characteristics expected in expert shiphandlers would be the capability to recognise when manoeuvring conditions could exceed acceptable limits. It was purposely used the word “shiphandlers” and not “pilots”, since the safety of a vessel is something that is certainly not limited only to pilots’ concern. Maritime regulations identify the Captain of every vessel as the ultimate responsible for the safety of crew and ship. Very often this responsibility is fulfilled through “preventive” decisions instead of “repairing” ones.

Our study wanted to investigate this particular aspect, including in the experimental design, conditions that marginally exceeded manoeuvring safety limits. Part of the research aimed to identify if participants, not only were aware of such violation, but also if they were able to predict the implications and consequences for shiphandling in those conditions. Paper II was dedicated to this exploration (see sections 4.1.2 and 6.2 of this thesis) and the results will be better discussed in chapter 5.

Another element characterising expertise, as highlighted in the literature review presented in this chapter, is the way experts make their decision in the field. In this thesis it was decided to explore more specifically what is the process of data collection (in the simulator) that pilots followed to support their decision making. This topic was explored in paper IV (see sections 4.1.4 and 6.4) and its implications, related also to the concept of situation awareness are further discussed in section 5.1.3.

### 2.1.9. Expertise and Cognition.

Cognition, as adopted in our theoretical model (see Figure 2 in section 1.2.), was indirectly referred when describing the process through which pilots gained and maintained situational awareness. This paragraph is dedicated to expand on the notion of cognition as thoroughly defined in the literature.

Expertise acquisition, according to the traditional Fitts and Posner cognitive model (1967), initially requires beginners to understand what a task requires, focusing on generating correct actions while avoiding gross mistakes (Feltovich et al., 2006). With practise, gross errors become increasingly rare and performance increases, requiring much less intensive focus. Further training and experience, will allow subjects to reach an acceptable level of performance. Individuals will start to adapt specifically to a learning domain, automating low level cognitive skills, showing more precision and less effort in the execution of a task.

As detailed in previous paragraphs, automaticity is a consequence of this refining process, where performers may partially lose the ability to control the execution of some actions. Intentional modifications and adjustments may at this stage become difficult. Performance gains stability, no further dramatic improvements are shown, but suitable results for everyday activities are ensured. The majority of individuals, follow the described process, moving through the “cognitive” and “associative” phases, until they reach the capability to perform virtually automatically, with a minimal amount of effort. What differentiates “experts” from “journeymen” (competent, yet less expert performers), seems to be the management of automaticity. Experts tend to remain in the “cognitive” and “associative” phases, while developing increasingly complex mental representations. In this way they maintain higher levels of control of their performance.

An example can be provided from early work in physics (M. T. H. Chi, Feltovich, et al., 1981) and medicine (Feltovich et al., 1984; Johnson et al., 1981). The expert’s strategy to solve presented problems was based on grouping them according to major physics principles, instead of salient objects contained in the problem statement itself. Experts exploit “Moderately Abstracted Conceptual Representations” (MACRs) (Zeitz, 1997), to more efficiently retrieve appropriate material from memory and to better integrate information selecting only what is important (Phelps & Shanteau, 1978). Conceptual representations guide the general approach to a problem (Schmidt, 1989; Voss et al., 1983) aiding productive analogical reasoning (Gentner, 1988) and providing alternatives (Patel, Arocha, & Kaufman, 1994). The nature of experts’ representations is functional and extends to entire activities or events. It is oriented to support planning, reasoning, monitoring, and evaluation (K. A. K. Ericsson, Walter, 2000), as showed in studies about fire fighters (G. Klein, 1999), and surgeons (Koschmann, LeBaron, Goodwin, & Feltovich, 2001).

Another aspect that plays an important role in the way experts gain and employ their knowledge is the phenomenon of Intuition. Gobet and Chassy(2009), reviewed Hubert Dreyfus' theory (H. Dreyfus, 1972; H. L. Dreyfus, Dreyfus, & Zadeh, 1987), where intuition is a signature of the holistic processing of the brain and the mind, and Herbert Simon theory (Chase & Simon, 1973; Herbert A Simon, 1989), where a simple mechanisms, based on pattern recognition, was sufficient for explaining intuition. Gobet and Chassy(2009), with their "template theory" attempt to account for empirical data linked to intuition. They introduce Templates that are schema-like structures that enable information to be encoded both rapidly and at a high level of representation. They argue that templates, in addition to pattern recognition (already present in the chunking theory) is the key mechanisms to explain the interaction between perception, attention, and learning. A further important addition in their theory, was linking chunks and templates to emotions. In addition, part of the theory was formally expressed as a computer program. The authors conclusion is that, while aspects of expert intuition can be characterized as holistic, the mechanisms that lead to them are local. Koziol, Budding et al. (2010), in a recent comprehensive review, offered an interesting framework where the most current cognition models are explained in relation to the underlying neuroanatomical structures.

In the model adopted in this study (figure 1), pilots' cognition was mainly mentioned in relation to the processes involved in the acquisition of situation awareness. Conducting a vessel requires a deep understanding of what it is happening on board and around and prevention is the keyword. Masses and dimensions involved do not allow gross mistakes: turning radiuses are in the order of hundreds of meters (if not km) and displacements (weigh of a vessel) are in the order of hundreds of thousands of tons. Consequences of an unclear understanding of manoeuvre developments are severe. Nevertheless the complexity of these operations cannot be worse than other industrial contexts (nuclear plants, air traffic controllers..). Similarly to other fields, also in pilotage cognitive strategies are required to simplify complexity in more manageable constellations of familiar and simpler scenarios. Pre-empting hydrodynamic effects of currents, exploiting predominant winds, expecting specific behaviours of local traffic (a ferry crossing the port at certain times of the day..), in the maritime world would be referred as "pilots' local knowledge". Vice versa, such local knowledge would build from those useful constellations that proved to ease pilots' job ("rules of thumb"). Pilots' expertise, among other elements, would also be formed by the amount of these rules and, most importantly, by their correct application in the specific shiphandling context.

In light of the above considerations, it may become clearer why planning a manoeuvre was adopted in this methodology. Predicting in advance, with the required precision, the strategy required to obtain a certain outcome in a shiphandling manoeuvre, implies to own those complex “chunks” of information or knowledge as well as their correct application in the provided context. In this work it was not specifically analysed the cognitive process that allowed pilots to generate a manoeuvring plan. The interest was more focused on the plan itself, considered as the direct outcome of pilots’ cognitive processes. Such outcome is what it is referred in Figure 2 as “mental models”. The following paragraph will help us to better describe how mental models have been introduced and described in the literature and how such concept becomes relevant for this research.

## 2.2. Mental Models

Mental models have been succinctly defined as “mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future states” ((Rouse & Morris, 1986), p. 351). They are generally used to describe a person’s representation of some physical system, but they can also be used to describe abstract dynamics or concepts. Mental models, seen as knowledge structures (Johnson-Laird, 1983), are formed of stored long-term static information that can be exploited to explain, interact and direct problem solving and their nature is different from the dynamically changeable situation awareness (Mica R. Endsley, 2006).

Problem solving in the real world indicate that when complex, novel, high-risk problems are presented, often also requiring a creative approach, people rely on mental models as a guide (Mumford et al., 2012). Fiol and Huff (1992), in their review, link mental models to maps. Like maps, mental models are virtual representations that are able to locate people in relation to their information environments. They provide a frame of reference for what is known and believed. They highlight some information and may fail to include other information, either because deemed less important, or because unknown. They hold the reasoning behind purposeful actions.

By exploiting this mental representations, specific aspects of performance can be improved, coordinated and adjusted. A well-developed mental model is based on aspects of the system that are the most relevant. Attention is directed according to this knowledge, to support slim and efficient perception processing, especially when a large amount of information is available. These models intend to integrate perceived elements in a meaningful ensemble, when single elements, approached one at a time, would not be able to convey the full picture. Provided with an accurate mental model, continuously confronted and tested against reality, individuals can project in the future actual system states in order to find the path to their goals.

It was previously identified how Chess players' capability to manipulate chess positions in long-term memory develops as a function of increased chess skill (K. Anders Ericsson & Kintsch, 1995; K. A. K. Ericsson, Walter, 2000). They engage in deliberate practise by analysing published games, playing through, to determine similarities in their moves and those chosen by masters. Through these efforts, less experienced chess players are able to experiment changes and improve their selection, widening the limits of their gaming mental models (N. Charness et al., 1996; N. Charness et al., 2005; K. Anders Ericsson et al., 1993).

Also much simpler tasks such as type writing, seem to imply mental planning. High-speed cameras were able to show how, finger movements toward the desired locations on the keyboard started well before the keys were struck. Typist were looking ahead in the text to prepare future keystrokes in advance (Timothy A. Salthouse, 1984). Tennis players, were able to anticipate where a ball was landing, exploiting subtle motion cues (A. M. Williams, Ward, Knowles, & Smeeton, 2002). These results strongly suggest that anticipation, mediated by cognitive mental models or representations, may be the major factor that explains superior speed of reactions in expert performers, rather than faster basic speed of their nervous system (Abernethy, 1991). Mental models can also be used to describe abstract dynamics or concepts as deductive reasoning and inference (Aronson, 1997), they could refer to individual or distributed cognitive processes among team members (Banks & Millward, 2000). Effective planning increases shared mental models, allowing team members to better perform during high workload conditions (Stout, Cannon-Bowers, Salas, & Milanovich, 1999).

At the end of paragraph 2.1.9 (Expertise and Cognition) it was introduced how, also in pilotage, "chunks" of knowledge need to be acquired and adopted in cognitive strategies. It was provided the example of "pilots' local knowledge" as the amount of contextualized information that is used to pre-empt, exploit or counteract known conditions and events that occur in the working environment. In this paragraph, the wider concept of mental models and their fundamental importance in providing guidance to plan and accomplish complex tasks is expanded further.

In our methodology, those plans had a pivotal role. Since it is reasonably challenging to measure internal cognitive processes, those plans were used as the overt manifestation of those processes. Paper II (see sections 4.1.2 and 6.2) illustrated how manoeuvring plans were considered the practical translation of pilots' mental models and, differently from cognitive processes, they could be accessed to obtain empirical measurements. As depicted in Figure 2, mental models were mentioned in two occasions in our model. They were initially created during the planning stage, where pilots analysed the provided manoeuvring conditions and activated the relevant knowledge, providing their strategy in the form of a plan. Plans, allowing empirical measurements, served as the basis against which the executions of manoeuvres could be compared. Mental models are then mentioned during the execution stage, when they provided the desired state at which pilot were aiming during the conduction of the manoeuvre. Comparing the measurements obtained from the plans against those obtained from the simulator, specific performance variables were obtained, as described in section 3.3.3 of this thesis. In section 5.1.1 the discussion on how those variables were related to the wider concept of expertise will be deepen.

## 2.3. Mental Workload

Previous sections have described how Expertise and Experts' characteristics varied according to the many fields of application, though maintaining certain communalities. One of these common elements shared among experts was their capability to generate accurate mental models able to represent a reliable depiction of the specific context they were operating in. Another element that was introduced in the previously described studies, was the relationship between the level of effort required to achieve certain outcomes and the level of expertise. The processes of perceiving, understanding, recalling and comparing with previous experiences, analysing, projecting, then acting with precision and effectiveness and then reassessing the outcome, puts a different "burden" on the shoulders of the subjects involved, depending on their level of expertise.

This "burden" might be otherwise referred to as Mental Workload. Despite interest in the topic for the last few decades (Huey & Wickens, 1993), there is still no clearly defined and universally accepted definition of mental workload (Cain, 2007). Workload, as a mental construct, is considered a variable (Gopher & Donchin, 1986), dependant on the mental demands imposed on operators by different tasks. The operator's workload will also depend on his experience in that particular context. Workload is thought to be multidimensional and multifaceted, difficult to be uniquely defined. Table 5 introduces only a few of the many definitions of mental workload that can be found in the literature:

Table 5. Few formal definitions of workload

| **Definition** | **Author** |
|---|---|
| *"Mental workload refers to the portion of operator information processing capacity or resources that is actually required to meet system demands."* | (Eggemeier et al., 1991) *p. 207* |
| *"... mental workload may be viewed as the difference between the capacities of the information processing system that are required for task performance to satisfy performance expectations and the capacity available at any given time."* | (Gopher & Donchin, 1986) *p. 41-3* |
| *"… the mental effort that the human operator devotes to control or supervision relative to his capacity to expend mental effort … workload is never greater than unity."* | (Curry, Jex, Levison, & Stassen, 1979) |
| *"... the cost of performing a task in terms of a reduction in the capacity to perform additional tasks that use the same processing resource."* | (Arthur F. Kramer, Sirevaag, & Braune, 1987) |
| *"... the relative capacity to respond, the emphasis is on predicting what the operator will be able to accomplish in the future."* | (Lysaght, Hill, Dick, Plamondon, & Linton, 1989) *p. 27* |

Workload can also be seen as an attempt to characterize performance in a task, relative to the operator's capability (Gopher & Braune, 1984). Gopher and Braune assume that workload is the effect of task demands on a single, undifferentiated pool of resources. In this view, multiple concurrent tasks will add together in their effect on the operator

(Gopher & Braune, 1984). Recent theories suggest that operator's resources are engaged differently and even independently, depending on the type of task (Hollands & Wickens, 1999; Jex, 1988). This approach was considered by Grier, Wickens et al (2008), when investigating the definition of an upper "red line" for task demands. In their study such red line was described as the amount of workload (as related to task demand) beyond which performance would drop below an "unacceptable" level. With the aim to predict operationally meaningful differences in performance within multi tasking settings, Wickens' model (2002) defined a four dimensions model, with separate resources able to address in parallel different types of task demands. Wickens' model described:

(a) three dimensions associated with different stages of processing;
(b) Different "codes" of processing (e.g., visual and language are "coded" differently for processing) and
(c) Different modalities (that indicates auditory perception uses different resources than does visual perception).

Subsequently a fourth dimension was added around visual channels to distinguish between focal and ambient vision. The value of the model is in "predicting relative differences in multi-tasking between different conditions" ((Christopher D Wickens, 2008) p.452). This makes unclear whether workload should be represented as a single scalar quantity or subdivided in several components. Of a different opinion is Wierwille ((1988) p. 318): he suggests that an operator would be fully engaged in a single task at a time, moving then from task to task, once completed.

Other constructs may interfere with the definition and the study of mental workload. One of those is stress. According to Gaillard (1993), both stress and workload depend on environmental demands related to the operator capability to cope with those demands. Gaillard, though, separates workload from emotions. He argues that workload is the result of the effort made by a higher, mental mechanism, similar to a meta controller, while affective factors play a complementary role to information processing, influencing the perception of workload. Gaillard, in this way, suggests a two dimensional model of cognitive energy mobilization. Colle and Reid (1999) assert that a one to one relationship between mental workload and other resources has yet to be defined in information processing theories. They describe mental workload as the average rate of mental work, detailing a procedure for defining a tasks' demand equivalence, but without identifying an appropriate time interval for a task in such an assessment. Huey and Wickens (1993) in their review, summarize many of the task factors affecting workload.

As identified above, a commonly accepted, formal definition of workload has not been adopted. Still, workload has been generally identified as a mental construct that reflects the mental "strain" when performing tasks under specific environmental and operational conditions, depending on the capability of the operator involved. Some authors, giving priority to one of these two main aspects of the problem (on one side the tasks with their demands, on the other side the operator with his capabilities), were able to choose among two different approaches:

*The task requirements* approach, where Mental Workload is viewed in terms of task requirements, and it is considered as an independent, external variable, with which the subjects have to cope more or less efficiently (P. Hancock & Chignell, 1986),

The *requirements resources interaction* approach where Mental Workload is defined in terms of interaction between human capabilities or resources and task requirements (P. Hancock & Chignell, 1986; Wieland-Eckelmann, 1992).

The *task requirements* approach was mainly adopted in occupational psychology and ergonomics, and subjects predominantly engaged in task design. Task requirements represent the stress and the strain would be the effort experienced by subjects, attempting to adapt to the demands (P. Hancock & Chignell, 1986). In this approach the attention is focused on how to design tasks in order to optimize their impact on subjects facing them.

The *requirements resources interaction* approach was developed within personality-environment fit-misfit theories, interested mainly in inter-individually different responses to identical physical and psychosocial conditions and requirements. The focus is on individual differences and on subjective responses to different conditions, in terms of fatigue, burnout or diseases (Gopher & Donchin, 1986; P. A. Hancock & Meshkati, 1988). In both cases Workload is derived from the complex interaction between the requirements of a task, the circumstances under which it is performed, and the skills, behaviours, and perceptions of the operator (Hart & Staveland, 1988).

Since workload cannot be directly observed, overt behaviour or measurement of psychological and physiological processes need to be gathered and used for inference (Casali & Wierwille, 1984). Not only it has not been defined a single representative measure of workload, but it is also difficult to suggest how many workload measures may be necessary or sufficient (Gopher & Donchin, 1986). Given the complexity and the many interactions that may affect the assessment of mental workload, as so far described, it is usually necessary to select a battery of measures in any experimental evaluation. In the following paragraphs, a list of different type of measures is provided and described.

### 2.3.1. Workload measures – Performance based

In this type of measurement what it is measured is the performance in the execution of a particular task. It is possible to refer then to primary-task and secondary-task measures. The primary task is the task whose workload is being investigated, whereas a secondary task is an additional task used to determine the operator's spare capacity. This approach assumes that performance directly measured on the second task is a function of the workload experienced to perform the first task (Navon & Gopher, 1979). With a complex task, though, the secondary task technique is much more difficult. It is important also to note that not all tasks interfere with each in the same way (Farmer, Berman, & Fletcher, 1986; C. Wickens, 1992). The appropriateness of particular performance measures is determined largely by the nature of the task. Common measures are reaction *time* and *accuracy*, in the form of percentage or proportion of errors, often coupled (Paul M Fitts, 1966). Another measure is the *root mean square (RMS) error*, calculated as the distance between actual and desired position. The use of RMS error more than the arithmetic mean, penalises inconsistency. More complex measures can be taken with weak, irregular signals, exploiting the Signal Detection Theory (Swets, 2014). In this thesis, it was not specifically utilised the type of measurements described in this paragraph. It was felt that this type of measurements would have excessively interfered with the execution of the proposed manoeuvres. The lack of feasibility of those measurements, if adopted in future field studies, was also considered.

### 2.3.2. Workload measures – Subjective Measures

The assumption in subjective measures or self-rating techniques is that the operator is the best evaluator of the mental effort he/she is experiencing while performing the required task. In the design of different measuring instruments, the objective was to develop rating scales able to provide a sensitive summary of workload variations within and between tasks, diagnostic of the sources of workload and relatively insensitive to the individual differences among subjects. Some examples are described below.

- Modified Cooper-Harper scale: The scale was originally developed for test pilots rating the handling qualities of aircrafts. Using a decision tree, it arrives at a value of 1–10.
- ISA: Instantaneous Self Assessment (ISA) it is a scale using which an operator can estimate perceived workload during real-time simulated or actual tasks. The perceived workload is rated on a scale from 1 (very low) to 5 (very high).
- RSME: RSME (Rating Scale Mental Effort) is a unidimensional scale developed in the Netherlands by Zijlstra (1985; 1993). Ratings of invested effort are indicated by a cross on a continuous line of 150 mm. (De Waard, 1996)
- Subjective Workload Assessment Technique (SWAT): This instrument includes scales for time load, mental effort load, and psychological stress load, each scale having three levels. Subjects are asked to rank the 27 possible combinations of levels on the

three scales before providing ratings for particular tasks or events (Reid, Shingledecker, Nygren, & Eggemeier, 1981).

- NASA Task Load Index (TLX): With this instrument, a final score is obtained after combining six rating scales using a weighted average, based on the ranking of these six scales (Hart & Staveland, 1988).

A strong point of the subjective scales is their high face value. Also, they have been thoroughly tested and validated. However, changes in subjective workload ratings can have several causes and do not necessarily reflect changes in the task demand load. Subjective rating tend to be more effective with rule or knowledge based tasks with conscious processing of data, less effective with highly practised skill based tasks (Van Westrenen, 1999).

In this research the NASA TLX was adopted and was administered at the end of each manoeuvre. This provided an overall evaluation of the workload experienced by pilots, in addition it helped to better define the nature of such workload, referring to the six subscales included in the questionnaire. To continuously measure the workload experienced throughout the manoeuvres, a Likert scale on seven levels was specifically built for this research (see Appendix 8 and section 3.3.3). The use of both these measurements was adopted in paper III (see sections 4.1.3 and 6.3) and the results obtained will be discussed in section 5.1.2 of this thesis.

### 2.3.3. Workload measures – Physiological

These measures are based on the premise that workload will induce bodily changes. In general, these measures are less convenient to use than performance and subjective measures, but they can provide useful additional information (Eric Farmer, 2003) Reliable techniques have been available to calculate physical workload. Oxygen consumption, heart-rate, and blood pressure can be used to estimate the subject's energy expenditure in performing the task. Blood pressure and heart-rate are known to be influenced by the effort required to perform a mental task. Other cardio-vascular parameters are more hidden.

As early as 1876 Mayer described variations in the heart-rate that were slower than the respiratory cycle (Penaz, 1978), and as early as 1963, Kalsbeek and Ettema found decreased heart-rate variability with increased task complexity (Kalsbeek & Ettema, 1963). These measures are related to the concept of arousal, a continuum that extends from deep sleep to a state of frantic excitement. An operator who is overloaded may experience increased arousal, manifested in changes such as increase in heart rate and skin conductance. Since workload is commonly considered to be a stressor, biochemical changes associated with stress, such as increase in cortisol excretion, are sometimes assessed in workload studies. The cardiovascular effects of mental tasks are also described to be similar to a defence reaction (G. Mulder, 1980).

The mechanism regulating arousal can be found in the part of the brain that controls homeostasis, the regulation of bodily functions under changing internal and external conditions. This mechanism can be found in the autonomic nervous system, a part of the peripheral nervous system. This part of the brain controls the heart, secreting glands, and involuntary muscles. The other part of interest is the central nervous system, the reticular formation, the nervous system within the skull and spinal column, that includes the brain, the brain stem and the spinal cord. Activity in these structures can be measured using various techniques and can provide highly valuable information about the mental activity involved in executing a task. The variables that can be obtained include brain activity, muscle tension, muscle tremor, pupil diameter, and eye blink rate.

Unfortunately, these techniques have limited applications outside a laboratory due to the high sensitivity for environmental noise or the need for complex and bulky measuring equipment. The relationship between mental activity and physiological measures and the relationship between physical activity and physiological measures make these measures very useful, however, by and large physiological techniques do not distinguish very well between physical effort and mental effort. Therefore the application requires a technique for filtering the effects caused by physical effort. This is most often achieved by minimizing physical effort or keeping it at a constant low level. A second and very serious drawback is the large amount of data which becomes available with physiological techniques.

Some of these physiological recording techniques are almost unobtrusive and make continuous recording during normal task execution possible. This is particularly true for the cardiovascular methods. The dynamic response of the different techniques varies widely. Techniques such as heart-rate respond in the order of seconds, heart-rate variability in tens of seconds, to several minutes for certain hormone secretions. The correct choice of tools depends largely on the situation at hand. Each single technique is only valid within its own particular set of constraints (Van Westrenen, 1999).

As described in section 3.3.3 of this thesis, some physiological measurements were specifically adopted. Paper III (see sections 4.1.3 and 6.3) explains how those measurements were collected and reports results obtained from their comparison with other workload subjective measures adopted in the study. Further discussion about the contribution of those measurements in the context of this research, will be provided in section 5.1.2.

In the following paragraphs a more detailed description of the different physiological measures is provided.

***Electrocardiogram (ECG) - Heart rate and HR Variability***

Several variables of the cardiovascular system are associated with workload, fatigue, arousal and stress, to which these variables all react differently. Fatigue is the result of a prolonged period of high levels of workload without periods of recuperation, most importantly, recuperation through sleep. Arousal is physiological readiness and stress is a physiological reaction to high levels of arousal for a prolonged period of time without recuperation. Blood pressure has been shown to be an indicator for arousal, stress and fatigue in aviation settings (Blix, Stromme, & Ursin, 1974; Nagle, Naughton, & Balke, 1966) but is not a good indicator for mental workload (Wierwille, 1979).

Heart rate is affected by physical workload, the readiness for bodily movements (Lysaght et al., 1989) and general arousal. Psychophysiology has contributed much to the development and understanding of the mental workload concept. The most studied and best understood response is heart rate, which is also one of the psychophysiological responses most frequently used for mental workload assessment (Jans Aasman, Mulder, & Mulder, 1987; Eggemeier et al., 1991; Hartman & McKenzie, 1979; Jiang et al., 1993; Jorna, 1992; Arthur F Kramer, 1991; L. Mulder, 1992; Nickel & Nachreiner, 2003; Roscoe, 1992; Van Steenis, Tulen, & Mulder, 1994).

The time series of the heartbeat is the basis of the HRV-technique. An ECG provides a continuous signal of the electrical activity of the heart. From this signal the occurrence of the R-peak is extracted and these peaks are used to calculate the frequency spectrum, transforming the heart-rate time series directly into a Fourier series based function, using a Discrete Fourier Transform. When the time series is transformed into the frequency domain, it is generally represented as a power-density function. This power-density function shows all the frequency components present in the time-signal, with their respective energy.

The heart-rate shows fluctuations of about 10% of the mean heart-rate. This is especially true under conditions of complete rest; with increased heart rates, this fluctuation generally decreases. These fluctuations are known as heart-rate variability or HRV. The magnitude of the HRV depends on the physical load, the mental load, and some unknown factors (Jans Aasman et al., 1987; Grossman, 1983; Kitamura, Murai, Hayashi, Fujita, & Maenaka, 2014; L. Mulder, 1992). When analysing the HRV, three components can be distinguished: one with a period of about 50s, one of about 10s, and one of about 3s. Since all components have a periodic characteristic, analysis is often done in the frequency domain where each of the components are clearly observable. The lowest frequency component is concentrated around 0.02 Hz and is related to slow regulatory processes such as thermo-regulatory activity (Rompelman, Coenen, & Kitney, 1977). The second component, which is related to the blood pressure regulating mechanism, focuses around 0.1Hz (0.07 ±0.14Hz is a practical spectrum, see (J Aasman, Wijers, Mulder, & Mulder, 1988). The third component, which results from the respiration mechanism, focuses around 0.3Hz. There is an increase of heart-rate during inspiration and a decrease during expiration. The HRV has shown sensitivity to mental load. Mental load decreases the HRV in both the middle and high frequency band but particularly the area around 0.1Hz (J Aasman et al., 1988).

Hence subsequent analysis in this thesis uses this band (see section 3.3.3 – Physiological Variables).

***Respiratory System***

Respiration is an automatic process, regulated by a respiratory centre in the hind-brain. Connections with the cerebral cortex make limited voluntary control possible. Respiration rate and ventilation have been found to be affected by emotional states, stress and arousal (Grossman, 1983; Thackray, 1969). Due to connections between the respiratory and the cardiovascular system the HRV shows a component around 0.40 Hz, the phenomenon known as respiratory sinus arrhythmia. The correlation between respiratory rate variability and mental load has been suggested, but extensive research has not been conducted. Respiratory measures are severely modulated by speech. This makes it inappropriate for use in a working environment where speech is essential.

This measurement was not adopted in this research, due to the explained difficulties to filter out speech artefacts.

***Electroencephalography (EEG)***

The EEG signal is a representation of brain's electrical activity recorded from electrodes placed on the scalp. It has been used to assess operators workload for many years in both laboratory (Berka et al., 2007; Gundel & Wilson, 1992; Lei, Welke, & Roetting, 2009) and applied settings (Kohlmorgen et al., 2007; G. F. Wilson, 2002). The EEG spectral components, for example, theta (4–8 Hz) and alpha (8–12 Hz), are used to determine activity levels during different cognitive activities. The majority of previous findings consistently indicate that increased workload leads to increased frontal theta (fro-theta) activity and decreased parietal alpha activity (Gevins et al., 1998; Gevins, Smith, McEvoy, & Yu, 1997; Gundel & Wilson, 1992; M. E. Smith, Gevins, Brown, Karnik, & Du, 2001; Sterman, Mann, Kaiser, & Suyenobu, 1994; Wu, Miwa, & Uchida, 2017; Yamamoto & Matsuoka, 1990). EEG spectrum modulation has also been introduced to investigate driver workload in various driving conditions (Brookhuis & De Waard, 1993; Hagemann, 2008; Kohlmorgen et al., 2007). Brookhuis and De Waard (1993) used an energy parameter ([theta + alpha] / beta) to measure participants activation during on-the-road driving experiments. In another study in the maritime industry, the analysis included both the Beta-1 (13-20 Hz) and the Beta-2 (20-36 Hz) bands, showing an increase in the Beta-2 band with increased mental work load (Koester, 2003a).

This measurement was adopted in this research (see section 3.3.3 – Physiological Variables).

***Pupillary response***

Pupillary response has been used in workload studies, the underlying rationale being that arousal will increase as a function of workload and so will do the pupil diameter. Although this measure has been used, it seems to be inapplicable in dynamic environments with changes of light conditions (Stone, Lee, Dennis, & Nettelbeck, 2004).

This measurement was adopted in this research, since automatically provided by the eye tracking devices adopted for gaze behavioural study (see section 3.3.3 – Physiological Variables). Recent studies have suggested the use of eye trackers to measure workload, even though there is still a lack of consensus on protocols and measures to be used to get meaningful results (Bjørneseth, Clarke, Dunlop, & Komandur, 2014; Di Nocera et al., 2016).

## 2.4. Situation Awareness

In the shiphandling expertise model presented in Figure 2 it can be noticed how situation awareness plays a prominent role. To better understand the relevance and the scope of that role, in this paragraph will be introduce the theoretical concept of situation awareness and some of its most renowned definitions and models as available in the literature. Further in the paragraph it will pointed out when a model or a definition will be specifically relevant to this research. A broader discussion on the relationship between expertise and situation awareness, will be held in section 5.1.3, in light of the results obtained with paper IV (reported in section 6.4) and highlighted in section 4.1.4 and 4.1.5 of this thesis.

As Dann reports in his comprehensive work (2012), the concept of "Situational Awareness" (SA) finds its roots in the aviation industry (Mica R. Endsley, 2006), where it started to be developed as early as World War II (Press, 1986). Since then, this theoretical construct has been adopted in several fields and several definitions have emerged accordingly. A summary and a classification of definitions can be found in Breton and Rousseau (2001). The classification divided the definitions into two main categories: SA defined as a State and SA defined as Process.

When defined as a *process*, situation awareness can be viewed just as an overarching concept encompassing several dynamic cognitive processes specifically related to event-driven, and multitask activities (Sarter & Woods, 1995).

When intended as a *state*, probably the most renown definition of SA is the one provided by Endsley: "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future"(M. R. Endsley, 1988).

In other words the distinction between "Process" and "State" mainly differentiates definitions that identify SA with the way it is achieved (based on underlying dynamic cognitive processes) from those definitions that see SA simply as the final product. Similarly, other differentiations of the definitions have been developed considering whether the approach is on the *Operator* or on the *Situation*. In the first approach the focus would be on the operator's mechanisms able to determine SA. The second approach, the focus would be on the environment or the working context where SA is achieved (Durso & Gronlund, 1999).

According to Breton and Rousseau (2001) another differentiation would be between *Prescriptive* and *Descriptive* models, where:

Prescriptive models are theory driven, characterized by processes and able to be translated into computational models. These models would allow numerical calculations based on modelled operators and environments. To provide some examples of these models: Zacharias and his colleagues proposed the SAMPLE model (Mulgund, Harper, Zacharias, & Menke, 2000; Zacharias, Miao, Illgen, Yara, & Siouris, 1996), essentially a diagnostic reasoning process. Shively, Brickner and Silbiger (1997) developed their model based on MIDAS architecture (Man-machine Integration Design and Analysis System). Starting from a set of elements connected to contextual information they defined SA as the weighted average over all those elements. McCarley, Wickens, Goh and Horrey (2002) were able to improve the previous model including projection. Despite of the differences, prescriptive models start from the definition of an ideal SA, as described by subject matter experts (SME). These models are based on a pre-established belief network, symbolising what the operator should consider in order to provide probability of actions at a given point in time.

On the other hand we have Descriptive Models, which attempt to capture actual SA process, considering operator's decision-making. The flexibility of those models, even though helpful to overcome operational setting constraints, reduce their possibility to be described through prescriptive algorithms, reducing their predictive power. Descriptive models though, providing a systematic description of SA and how it is achieved, are able to reflect operators decision making within a specific context, relating to a job or a role. Table 6 reports some definitions of Situation Awareness.

Table 6. Definitions of Situation Awareness (adapted from Dann (2012) Dominguez (1994) and Jeannot (2000))

| Definition | Author(s) |
|---|---|
| *SA is the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and their projection of their status into the near future.* | *M. Endsley* |
| *SA is "externally directed consciousness" with SA being "the invariant in the agent-environment system that generates the momentary knowledge and behaviour required to attain the goals specified by an arbiter of performance in the environment* | *Smith and Hancock* |
| *SA is the conscious dynamic reflection on the situation by an individual. It provides dynamic orientation to the situation, the opportunity to reflect not only on the past, present and future but the potential features of the situation. This dynamic reflection contains logical-conceptual, imaginative, conscious and unconscious components, which enables individuals to develop mental models of external events.* | *Bedny and Meister* |
| *SA is the "pilots internal model of the world around him at any point in time".* | *M. Endlsey* |
| *SA is the continuous extraction of information about a dynamic system or environment, the integration of this information with previously acquired knowledge to form a coherent mental picture, and the use of that picture in directing further perception of, anticipation of, and attention to future events. This definition is closely allied with Endsley's definition.* | *C. Wickens* |
| *SA is an individual's continuous extraction of environmental information and integration of this information with previous knowledge to form a coherent mental picture, and use that picture in directing future perception and anticipating future events.* | *Dominguez* |
| *SA is the perception of the elements in the environment within a volume of time and space, comprehension of their meaning and the projection of their status in the near future. This also means the continuous extraction of environmental information and the integration of this information with previous knowledge to form a coherent mental picture and the use of that picture in directing further perception and anticipating future events. SA is established by a continuous comparison between anticipation (predicted state of the system) and environmental input (actual state of the system). The AIM Human Resources Unit, Eurocontrol (Jeannot, 2000) adopted this definition and it can be considered to be a grouping of the definitions of Endsley (1988) and Dominguez (1994).* | *Jeannot* |
| *SA is the perception of reactions to a set of changing events. This definition emphasises the affordances in the situation and views a person's understanding in terms of what can be done (even if it is only to gather more data) instead of merely the recalled stimuli.* | *Klein* |
| *SA is the combining of new information with existing knowledge in working memory and the development of a composite picture of the situation along with projections of future status and subsequent decisions as to which course of action to take.* | *M. Fracker* |

Despite the differences, the definitions listed in Table 6 may have as a common denominator an individual's dynamic awareness of the external situation (Salmon et al., 2009). With this in mind, we can refer back to paragraph 2.1., and understand how expertise gained in a particular domain allows to develop and maintain situation awareness despite high volumes of information and system complexity, whereas a novice would be considerably overloaded, hampered by both limited attention and limited working memory. Novices would not grasp the importance of each piece of information, overseeing or missing important contextual elements. Another important element, shared by the several definitions, is that the development of situation awareness cannot be passive. Individuals drive the selection of information relevant to their domain. Experience may allow individuals to develop a reasonable level of automaticity. Low situation awareness caused by cognitive automaticity could negatively affect performance, especially outside of the learned routine. Experts in various domains apply different strategies trying to avoid the deleterious effects of automaticity. On the other hand, automaticity, offloading attentional demands, may be an important prerequisite for developing high levels of situation awareness. When very novel situations are encountered, and expert's mental model may be incomplete, there is the risk to try to apply an inappropriate schema. Even with a manual or psychomotor task, apparently not much related to a cognitive construct such as situation awareness, a correlation was found, presumably due to limited attention issues (Mica R Endsley & Bolstad, 1994; O'Hare, 1997).

Situation awareness may not be possible (depending on how the term is defined) as long as an individual must concentrate on the performance of the physical tasks involved. In army operations, for example, to develop situation awareness, army officers have traditionally employed numerous techniques. Studies using the situation awareness global assessment technique (SAGAT), conducted to determine differences in situation awareness between inexperienced and experienced army platoon leaders, showed that the more experienced group had higher SAGAT scores (Strater, Endsley, Pleban, & Matthews, 2001). Experience also affects situation awareness, shifting platoon leaders' focus from concentrating mainly on friendly disposition, to take more into account enemy disposition (Shattuck, Graham, Merlo, & Hah, 2000). Results were confirmed by additional studies, with inexperienced platoon leaders having many difficulties with forming good situation awareness (Strater, Jones, & Endsley, 2001). Situation awareness research was carried out in vehicle driving demonstrated a significant negative correlation between the time needed for predicting an hazard in a simulated driving task and drivers' reported accident rates (Currib, 1969; FP McKenna & Crick, 1994; F McKenna & Crick, 1997; Pelz & Krupat, 1974). This result remained consistent even after age and miles driven were controlled for (FP McKenna & Crick, 1994; Quimby, 1987). Experienced drivers detect more hazards and react faster (FP McKenna & Crick, 1994). Novice drivers pay attention to different areas than experienced drivers (Horswill & McKenna, 2004), using less effective scanning patterns (Mourant & Rockwell, 1972) when looking for hazards (Underwood, Chapman, Bowden, & Crundall, 2002). Hazard awareness and prediction require significant cognitive resources, allowing very low levels of automaticity. Exposure to a concurrent memory task has shown a significant impairment in driving performance (FP McKenna & Farrand, 1999). Similarly to the aviation and military domains, less experienced drivers are able to acknowledge less information, while expert drivers, exploiting automaticity on physical tasks and owning effective mental models, better direct information search and interpretation. Experts spend considerable effort at the task of situation assessment, actively projecting and planning for contingencies. Maintaining situation awareness and comparing gained critical cues with owned mental representations or models, experts can instantly recognize known classes of situations (Mica R Endsley & Bolstad, 1994). This mechanism is referred as "recognition-primed decision making" (as discussed in the earlier section 2.1.8 on expertise and decision making) (Gary A Klein et al., 1986) or "big switch," (Schank & Colby, 1973).

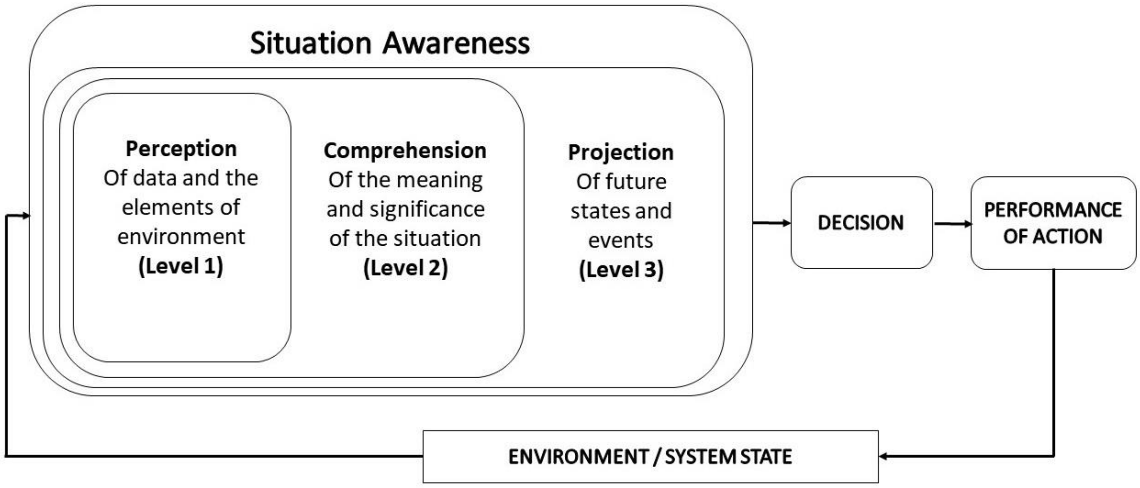
For the purpose of this thesis, three main theoretical approaches most adopted in the literature will be briefly described and compared:

1. Endsley's (1995) Three Level model:
   This model is based on an information processing approach. It describes SA as a product of the knowledge related outcomes of the three levels. It assumes that SA is separated from the processes used to achieve it (Woods & Sarter, 2010);
2. Smith and Hancock (1995) Perceptual Cycle model:
   This model postulates SA as an active interaction between humans and their environment. In this model it is the context of the interactions that defines SA, involving both the cognitive processes and the product of SA, continually updated (N. Stanton, Salmon, & Rafferty, 2013);
3. Bedny and Meister (1999) Activity Theory model:
   This model proposes SA as only one of many components of reflective-orientational activity. This model postulates a conscious dynamic reflection process which guides orientating attention and activity to the physical situation (Woods & Sarter, 2010).

**2.4.1. Endsley's Three Levels Model**

The three level model of SA is based on three linear levels as defined by Endsley's (1988): "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future". The model is based on an information processing approach. SA is considered a state of knowledge and is separated from the processes used to achieve it (N. Stanton et al., 2013). The model describes human information processes using psychological constructs such as attention and short term memory (Uhlarik & Comerford, 2002).

Figure 5. Endsley Three Level Model




Within this framework:

1. Level 1 of SA is the perception of the "status, attributes and dynamics of the relevant elements in the environment" (M.R. Endsley, 1995). The "perception of these cues is fundamental" as "without basic perception of important information, the odds of forming an incorrect picture of the situation increases dramatically" (M. R. Endsley & D. Garland, 2000).
2. Level 2 of SA is the comprehension of the current situation as the understanding of the information retained in working (Salas, Prince, Baker, & Shrestha, 1995) and long term memory (Christopher D Wickens, Gordon, Liu, & Lee, 1998).
3. Level 3 of SA is the projection of the actual situation into a future status as future actions of the elements in the environment. It is achieved through a knowledge of the status and dynamics of both level 1 and 2 of SA and once the comprehension of the activated schemata has been completed, the individuals are guided in their projection of the future status as well as their selection of future actions (C. A. Bolstad & Hess, 1995; M.R. Endsley, 1995).

The knowledge achieved at level 3 should allow for timely decision making, however this level has a demanding nature due the complexity of the underlying cognitive constructs (Christopher D Wickens et al., 1998). At this level, not only the individual has to be fully aware about what is going on, but needs also to be capable to accurately project such state in the future (M.R. Endsley, 1999). This level of situational awareness is particularly important for marine pilots. Although ships are moving slowly in comparison to other forms of transportation, the size and hydrodynamic properties of ships require the pilot to be able to correctly and timely project in the future: a bulk-carrier of a displacement of 110000 tonnes takes more than 4 Km to come to a complete stop.

### 2.4.2. Smith's and Hancock's Perceptual Cycle Model

The Perceptual Cycle model of SA was developed by Smith and Hancock (1995). This takes a more ecological and perhaps dynamic approach stressing the mutual and cyclical interaction between individual and environment. The model views SA as "a generative process of knowledge creation and informed action" (Salmon et al., 2008). The implied process is cyclical. This model is based upon Niesser's Perceptual Cycle (1976), where individuals use schemata to interact with the world and modify those schemata, based on previous iterative interactions. This model of SA (Uhlarik & Comerford, 2002) consists of three elements:

- The object
  Available information in the external environment;
- The schemata
  Internal knowledge developed through training and experience, and stored in long term memory;
- Exploration
  Process of active search of the environment conducted by the observer.

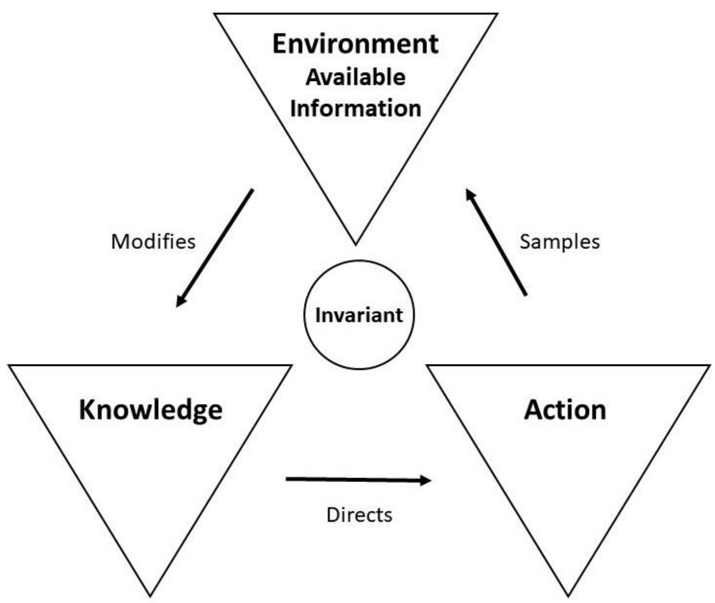
Smith's and Hancock's (1995) model refers to a perceptual cycle that identifies SA with internally held schemata (knowledge). Those schemata direct a person's interaction with the world. The outcome of the interaction, in turn, modifies the original schemata. Due to this modification, schemata will direct further interaction with the world. This process is cyclical and continuous. SA is the dynamic result of those reiterations of interactions of the person with the world. In this model, the only invariant is the capacity of the individual to adapt to the environment. The individual becomes the connecting link between information, knowledge and action.

The process, as described more in details in Figure 6, would be:

The invariant codifies the information available in the environment. Such information will then modify the knowledge that the agent requires to assess the situation. The invariant modulates the action that the knowledge will direct to attain goals. The invariant will sample the actions that will act on the environment. New information will be then available to restart the cycle.

Schemata, or internally held models, would already contain information regarding certain situations. These mental models facilitate the anticipation of situational events directing the person's attention to cues in the environment and directing their eventual course of action. The agent (individual) then carries out checks to confirm that the evolving situation conforms to his/her expectation (K. Smith & Hancock, 1995). Any unexpected event prompts further search and exploration and in turn modifies the individuals' existing model (N. Stanton et al., 2013).

Figure 6. Niesser's Perceptual Cycle




The conduction of a vessel can provide suitable examples. The whole position monitoring process can be seen as an application of the perceptual cycle. Pilots explore the available information looking at instruments, such as radars and chart displays, looking outside of bridge windows, observing navigational aids and landscape features (environment available information). They will select the most relevant information and assess ship position (knowledge). If the vessel is not in the expected position, they will decide what is the most suitable order that has to be given to correct the situation (action). Once the order is executed, they will start to explore the environment again to find the information that will confirm or not if their action was effective. Then the cycle will start again.

In the Perceptual Cycle model, at every cycle, knowledge is modified (or confirmed) and then directs further action (or not). The action that will interact with the environment has to be the result of a decision making process. So what are the elements that take part into this decision? Smith and Hancock (1995) underlined the importance of person's internally held models. As detailed in paragraph 2.2, these mental models contain information relevant to a certain situation. Mental models are able to anticipate information not yet perceived, they can integrate perceived elements in a meaningful ensemble, conveying the full picture. Mental models become the element against which the perceived reality is compared. Discrepancies between mental models and reality will direct remedial actions and / or further investigation of the information available. Unexpected events would prompt further exploration, in turn, modifying existing mental models (N. Stanton et al., 2013).

This model was specifically adopted in the study presented in paper IV (see section 6.4 and 6.5). Results obtained and summarised in section 4.1.4 and 4.1.5, were able to provide an example of how, according to the theoretical model, pilots' interaction with the shiphandling environment (in the form of exploration and actions), changed and adapted to manoeuvring conditions. A detailed discussion of the results in the context of this thesis is provided in section 5.1.3.

### 2.4.3. Bedny and Meister's Activity Theory Model

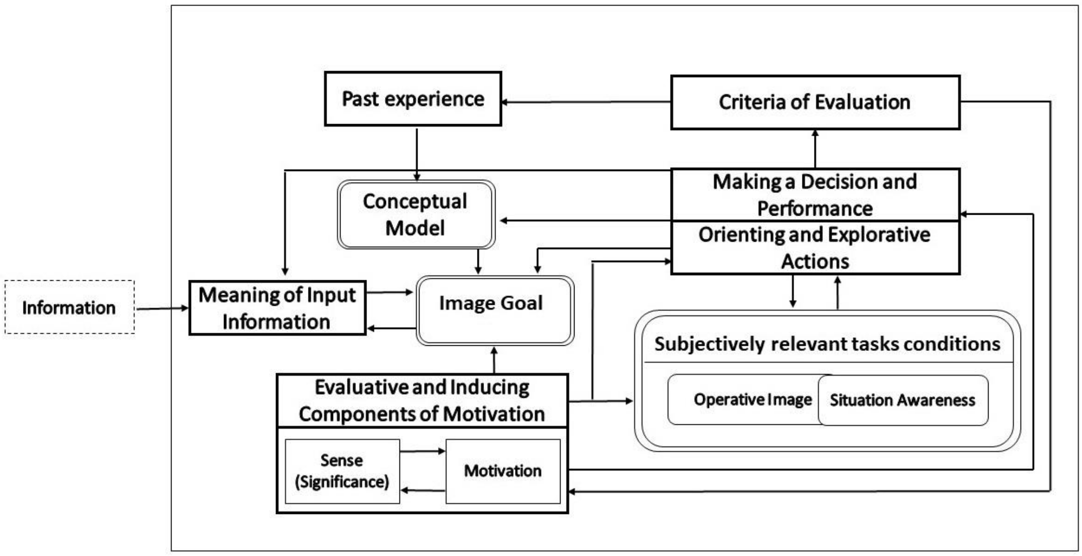
In Bedny's and Meister's (1999) Activity Theory Model the basic structural components of activity are goals. Goals are "an ideal image or a logical representation of future results of activity" ((Bedny & Meister, 1999), pag. 2). Goals can be internally originated or externally provided. Goals direct activity towards a desired outcome. The methods of activity (or actions) allow the achievement of these goals. The psychological distance between the goals and the current situation is evaluated from a personal significance point of view. Such distance, but also the significance of the failing to attaining that goal, motivates an individual to conduct activity towards achieving the goal.

The activity comprises three components (Bedny & Meister, 1999):

1. The orientational component involves the development of a subjective internal representation of the current situation. It provides a meaningful interpretation of reality and anticipation of future states.
2. The executive component entails the transformation of the situation, aiming towards a desired goal via decision-making and performance of action.
3. The evaluative component involves the assessment of the situation. Such feedback influences executive and orientational components.

Bedny and Meister's (1999) model is shown in Figure 7

Figure 7. Bedny and Meister's Activity Theory Model




Each model so far proposed has its own strengths and weaknesses as thoroughly argued by Dann in his work (2012).

Referring back to section 1.2 (Research Definitions) Figure 2 shows how the theoretical model adopted in this research emphasizes the cyclical nature of pilotage. For this reason, even though elements of Endsley's model will be mentioned and incorporated, this research will mainly refer to the Smith and Hancock's Perception Cycle Model.

## 2.5. The theoretical approach of the publications

Chapter 2 presented the theoretical framework underpinning the five papers of this research. In particular, Table 7 identifies the nature of the work undertaken and the aim of each paper in relation to this project, specifying the theoretical approach and framework, the underlying theories, as well as the way the theories have been used and/or further developed.

Thus, as indicated in Table 7, the theories presented in this chapter, have guided the research work undertaken in this study but have been further developed in order to adapt them to the particular context of the shiphandling expertise. This theoretical approach underpins the methodological choices made, as will be further explained in the next chapter. Thus, as illustrated below in Figure 8, the third chapter of the thesis will focus on the various research methods and materials used in this study. In particular, it will explain how data was collected and which analysis techniques were used in order to support the theoretical approaches described in this chapter as well as to address the research gaps and the research questions formulated in Chapter 1.

Figure 8. Progress Tracker – Moving to Chapter 3

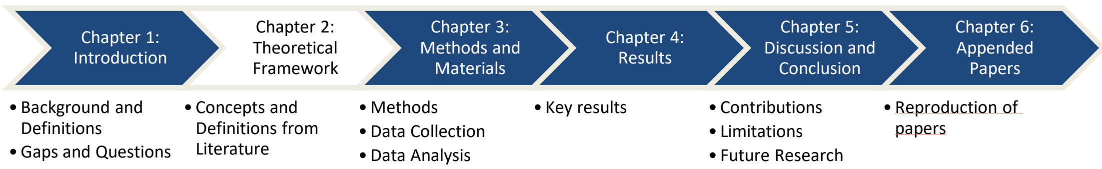


Table 7. Publications and their theoretical framework of reference

| Paper | Nature of the paper | Aim of the paper in relation to the project | Underlying theory | Theory use and/or development |
|---|---|---|---|---|
| I | Introductory | Introduction of the methodology and the theoretical framework adopted | Expertise<br>Mental models<br>Situation awareness | The paper introduced the proposed methodology and presented the adopted model within the more general theoretical framework of Expertise, Mental Models and Situation Awareness. The study did not provide analytical results but described the methods and devices that were adopted in the research. |
| II | Research paper | Quantification and evaluation of Mental Models and performance | Mental models<br>Expertise | The paper adopted the concept of mental models to: (1) apply it to marine pilotage, with reference to the already established practice to prepare passage plans; (2) explained how an analytical approach was adopted to translate a theoretical construct into a measurable and so comparable entity (3); demonstrated how mental models could be compared to execution to obtain performance. |
| III | Research paper | Direct and indirect quantification of mental workload | Mental workload<br>Expertise | The paper explored the use of several concurrent physiological and subjective measurements, known from literature to relate to mental workload. The aim was to identify the most suitable to be adopted in a shiphandling environment, able to provide an indirect and early indication of reached critical limits. The paper demonstrated how a simulated environment could be effective to elicit significant responses. |
| IV | Research paper (2 Versions) | Analysis of behavioural patterns to quantify situational awareness | Situation awareness<br>Expertise | The paper (provided in 2 versions) explained how the use of eye trackers can significantly help to identify and quantify, sequences of behavioural markers related to different cognitive processes involved in situational awareness (as defined by the Smith and Hancock's Cycle Model). This paper offered a first example of an analytical approach able to unpack expert behaviours in a shiphandling context. |

# 3. METHODS AND MATERIALS

This chapter focuses on the methods selected for this research, and explains how and why data was collected. It also explores the relationship between the theories presented in the previous chapter and the need for quantitative data collection and analysis.

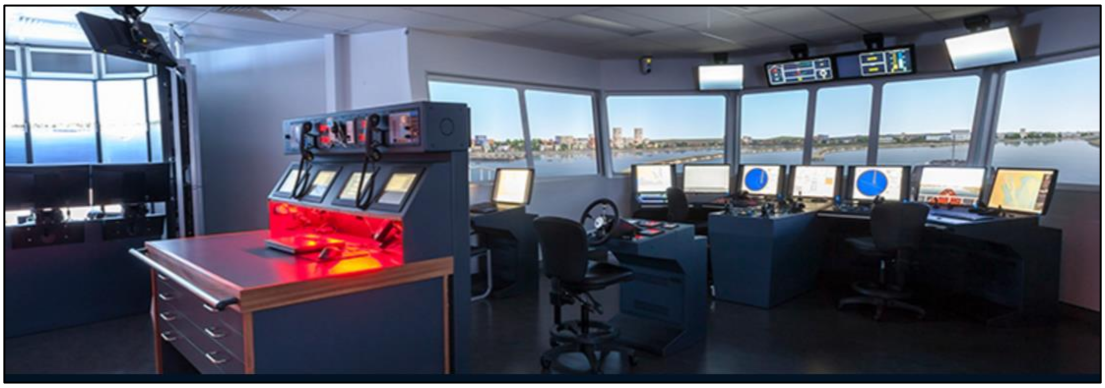
## 3.1. Participants

Participants included 10 marine pilots, employed by an Australian pilot company. They answered to an internal advertisement letter (Appendix 1) emailed within the Company after an initial presentation of the project to the Company management. A more detailed Information Sheet (Appendix 2) was provided to those interested and finally a Consent Form (Appendix 3) was signed by those willing to participate. The participants were all males in good health, as required by national professional medical standards set by the Australian Maritime Safety Authority (2010). Those professional standards imply a biannual (or yearly, if over 55) full medical check, which includes hearing and eyesight checks. An Analysis of Variance (ANOVA) for age and service confirmed no significant difference between the participants and the rest of the pilot group working for the same company. All the pilots involved in the research had more than ten years of previous experience as qualified pilot.

## 3.2. Setting – Simulator

To complete the data collection presented in this research, the Maritime Safety Queensland, Full Mission Bridge Simulator in Brisbane was used (Smartship® Simulator www.smartshipaustralia.com.au). This Simulator included advanced features such as a 16m diameter screen with 360 degree field of view (FOV). In addition, the Bridge was fully equipped with original bridge consoles featuring real navigation equipment and, in particular, NACOS 65-5 (a command and control, Integrated Navigational System) manufactured by SAM Electronics. The simulator software and hardware were provided by FORCE Technology® Denmark (www.forcetechnology.com). Even though the facility provided five different bridges, a Full Mission Bridge, classified as Class A (NAV) according to the standards issued by the classification society Det Norske Veritas Germanischer Lloyd (2014), was adopted to carry out this study. Those standards require that such Simulator should be capable of simulating a total shipboard bridge operation situation, including the capability for advanced manoeuvring in restricted waterways.

Figure 9. Bridge 1 at Smartship Simulator



## 3.3. Experimental Design

### 3.3.1. Assumptions

The main aim of the research was to collect objective measurements that could be able to quantify in a numerical form, pilots' expertise. Specific constructs such as situation awareness, mental models and workload, known from the literature to be related to expertise (as detailed in Chapter 2) were specifically targeted. The simulator was used to create different scenarios (independent variables) with the aim to explore differences in measurements (independent variables) related to those constructs. The present research did not aim at identifying individual differences among participants, rather focused on group changes of behaviours, performance outcomes, physiological reactions etc.. as induced by the different scenarios. Based on actual industry standards, the assumption was that the selected group of participants was a group of "experts". Our aim was to study how this group of experts would have responded to the different experimental conditions.

### 3.3.2. Independent Variables

The simulator was used to manipulate the different experimental conditions chosen for this study. The experimental conditions were four shiphandling manoeuvres carried out from the Bridge of a commercial vessel. Each manoeuvre included the whole process necessary to transfer the ship from a defined initial position to a berth within constrained port waters with the use of own and/or external means of propulsion (i.e. tug boats to assist, when allowed). In each of those manoeuvres, three main factors were controlled. These three main factors were: (a) "port familiarity" (from now on referred as "port"), (b) "difficulty", and (c) "phase". Table 8 provides a schematic of the three factors adopted in the experimental conditions.

Table 8. Experimental Design – List of factors per manoeuvre.

| | ***Manoeuvre "n" of 4*** | | | | |
|---|---|---|---|---|---|
| | ***(b) Level of Difficulty*** – Easy "0" or Difficult "1" | | | | |
| ***(a) Port Familiarity*** - Home "0" or Foreign port "1" | ***(c) Phase*** Baseline Pre "0" | ***(c) Phase*** Approach "1" | ***(c) Phase*** Swing "2" | ***(c) Phase*** Closing "3" | ***(c) Phase*** Baseline Post "4" |

***Port***

The first factor, "port", took into account whether the manoeuvres were conducted in the participant pilots' homeport (the port where they were regularly working) or in a foreign port. The foreign port was a virtual port only present in the simulator software. This port was chosen to avoid any possibility of a learning effect associated with previous manoeuvring experience the subjects may have owned and to provide support for methodology reliability. To test such reliability, in fact, spatial constraints related to port dimensions were purposely maintained identical, modifying the virtual port in order to match homeport dimensions as summarised in Table 9.

Table 9. Proportions between vessels and port dimensions.

| **Ship** | **LOA (m)** | **Ratio between Ships** | **Breadth (m)** | **Disp (ton)** |
|---|---|---|---|---|
| **Torm Laura (diff Lvl 0)** | *183* | *0.7* | *32* | *54925* |
| **Arcturus (diff Lvl 1)** | *269* | *1.45* | *48* | *143200* |
| **Ratio** | **Torm LOA** | **Torm Breadth** | **Arcturus LOA** | **Arcturus Breadth** |
| **Basin diameter (470 m)** | *2.6* | *14.7* | *1.7* | *9.8* |
| **Channel width (300 m)** | *1.6* | *9.3* | *1.1* | *6.2* |

The pilots' homeport in the tables and graphs presented is coded "1", while the virtual port is coded "2".

### *Difficulty*

The second factor was the shiphandling level of "difficulty". To control the level of difficulty, specific manoeuvres' parameters were altered as summarised in Table 10. The easy level is coded "1" while the difficulty level is coded "2".

Table 10. Levels of Difficulty – Adopted in both Ports.

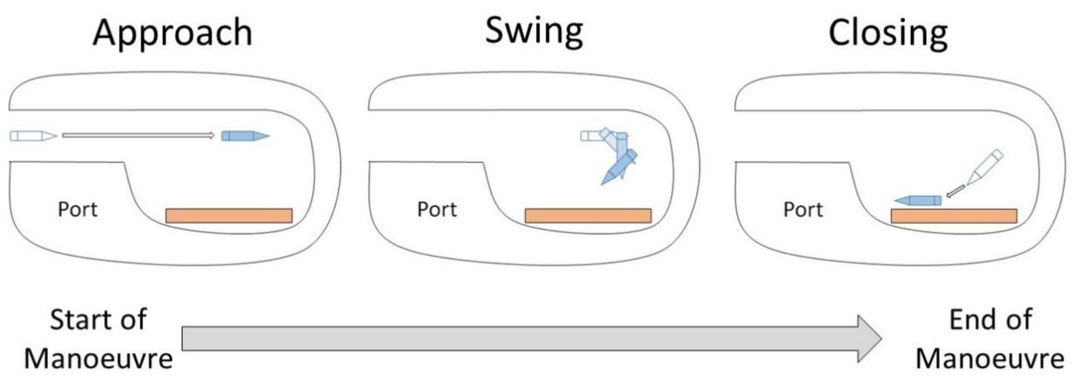
| | **Pier - Spatial constraints** | **Environmental conditions and forces** | **Vessel characteristics** | **Tugs** | **Interactions with traffic** | **VTS Comms [(1)]** |
|---|---|---|---|---|---|---|
| **Level 1**<br>**Easy** | Big Swing Basin<br>(3 times Vessel LOA [(4)]) | Current: 0.7 Knt<br>Wind: 15 Knt<br>Good Visibility | Single Controllable Pitch Propeller [(2)]<br>Bow Thruster [(3)] | None | 1 Interacting but not Interfering vessel | General Info<br>No reporting Points |
| **Level 2**<br>**Difficult** | Small Swing Basin [(5)]<br>(1,5 times Vessel LOA) | Current: 2 Knt<br>Wind: 25 Knt<br>Poor to no Visibility - Heavy Rain | Single Fixed Pitch Propeller<br>No Thrusters | As required by Pilot | 1 Interacting 1 Interfering vessels | General Info and Traffic Advice<br>Reporting Points |
| **Notes** | (1) Vessel Traffic Management station present in a port and managing ships via radio communications;<br>(2) Propeller capable to change the water thrust direction changing the angle of the blades instead of direction of rotation;<br>(3) A thruster is a propeller positioned perpendicular to the ship keel axis. Placed on the bow or on the stern, induces transversal / angular motion;<br>(4) Length Over All, maximum length of a vessel;<br>(5) Wider area, within constrained waters, where ships have sufficient room to rotate and revert their direction. | | | | | |

Level 1 reproduced a comparable level of difficulty of routine operations, a condition very similar to an ordinary job, regularly carried out in daily activities. Level 2 aimed to engage pilots with a scenario slightly exceeding the safety limits established in the pilots' homeport.

### *Phase*

The third factor was the "phase" into which every manoeuvre could be divided. Each manoeuvre required pilots to complete a mooring using the side of the ship opposite to the berth position on commencement of the exercise. This implied that for each manoeuvre the ship had to swing (rotate 180°) before she could be moored. Each manoeuvre therefore had to be developed through three main phases. These three phases were: the “approach” (from the initial position until the start of the swing), the “swing” (from the start of the swing until the rotation was completed and stabilised), and the “closing” (from the end of the swing until a defined distance from the berth). Figure 10 shows a schematic, summarising the three manoeuvring phases as they develop throughout each manoeuvre.

Figure 10. Schematic illustrating the three phases of each manoeuvre.




Two additional phases were considered for the purpose of collecting baseline data for some physiological variables, before and after the execution of every manoeuvre.

In summary, four different manoeuvres were conducted by each pilot (see Table 8). Two of the four manoeuvres were conducted in pilots’ homeport (one manoeuvre for each level of difficulty). Two manoeuvres were conducted in the foreign port (one manoeuvre for each level of difficulty). Exactly the same four manoeuvres were conducted by each participant, in random order, to mitigate a possible learning effect. The manoeuvres were presented to pilots before being performed in the simulator, since every participant was required to provide a plan, such as the one normally discussed by pilots and captains before a ship enters into a port (Wild & Constable, 2013). Before the experimental manoeuvres were conducted in the simulator, pilots were required to perform a very simple mooring with a vessel different from those used in the experimental manoeuvres. This first manoeuvre was used as a familiarisation manoeuvre.

### 3.3.3. Dependent Variables

Several dependant variables where chosen to be considered in this research. Below is provided only a brief summary and description of the variables as published in dedicated papers. A more detailed explanation on how these independent variables were obtained and processed will be extensively provided in section 3.5. of this thesis.

***Performance Variables***

The description of the following variables and the recorded outcomes have been discussed in the published paper “A Comparison of Marine Pilots’ Planning and Manoeuvring Skills: Uncovering Mental Models to Assess Shiphandling and Explore Expertise” (Luca Orlandi, Brooks, & Bowles, 2015), reported in section 6.2 of this thesis. Key results are provided in section 4.1.2.

- XTD – Cross Track Distance: Distance from the intended track as per positions obtained from the planning charts and the ship track provided by the simulator;
- SpdEst – Speed Estimation: Difference between the intended speed over the ground (SOG) as per DMP (expressed in knots) and the recorded speed provided by the simulator;
- EngEst – Engine Power Estimation: Difference between the absolute value of the intended use of engine power as per DMP (expressed in percentage) and the absolute value of the recorded engine power provided by the simulator.
- ThrEst – Bow Thruster Power Estimation: Difference between the absolute value of the intended bow thruster power (expressed in percentage) as per DMP and the recorded absolute value of the bow thruster power provided by the simulator (when applicable);
- Tug(n)Est – Tug Force Estimation: Difference between the absolute value of the forecasted tug’s bollard pull as per DMP (expressed in percentage, based on the maximum bollard pull that tugs could provide) and the recorded absolute value of the tug’s bollard pull provided by the simulator (when applicable, with (n) differentiating each tug used).

***Workload Variables***

The description of the following variables and the recorded outcomes have been discussed in the paper “Measuring mental workload and physiological reactions in marine pilots: building bridges towards redlines of performance” (Luca Orlandi & Brooks, 2018), at the moment under review and included in section 6.3 of this thesis. Key results are provided in section 4.1.3.

- NASA TLX – NASA-TLX (Hart & Staveland, 1988) is a multidimensional scale for which the overall mental workload score is obtained as a function of 6 subscales: 1. Mental Demand (MD), 2. Physical Demand (PD), 3. Temporal Demand (TD), 4. Own Performance (OP), 5. Effort (EF), 6. Frustration Level (FR);
- Self Assessment Likert Scale –A Likert scale was built specifically for this study, to verbally report subjective levels of "involvement or workload". The scale adopted, used 7 different levels of "experienced difficulty of the exercise" (see Appendix 8), meaning the personal level of workload experienced or effort necessary, in order to be able to manage the contingent situation;
- ECG – HR and ECG – LF/HF. Subjects' ECG was continuously recorded during the manoeuvres. From the ECG signal it was possible to obtain two specific measurements:
    - The Heart Rate, recorded as the number of heart beats per minute;
    - The LF/HF index, obtained as the ratio between the power spectral density calculated in the low frequencies and the power spectral density calculated in the high frequencies of the Heart Variability function;
- Pupil Dilation – The pupil dilation was obtained directly from the eye trackers worn by pilots during the manoeuvres and it was recorded as the diameter measured in pixels of the eye iris;
- EEG – Bands Beta 1 and Beta 2 PSD – Subjects' EEG was continuously recorded during the manoeuvres. From the EEG it was possible to obtain two specific measurements:
    - Beta 1 psd, obtained as the power spectral density calculated in the band Beta 1 (13 – 20 Hz) of the EEG signals recorded.
    - Beta 2 psd, obtained as the power spectral density calculated in the band Beta 2 (20 – 36 Hz) of the EEG signals recorded.

***Behavioural Variables***

The description of the following variables and the recorded outcomes have been discussed in the paper "Interpreting changes in marine pilots' perceptual cycle through gaze detection." (L Orlandi, Brooks, Wood, & Black, 201x), at the moment under review and included in section 6.4 of this thesis. In section 6.5 an unpublished version of the paper is also provided, since this version included additional analysis and results not reported in the submitted paper. Key results for both versions are reported in section 4.1.4. In Appendix 9 a further summary of the following variables is provided.

- Visual Position Check – This variable reports the frequency at which the pilots gaze was directed towards an external object on a certain relative direction and then moved to another object, approximately at 90 degrees of bearing from the previous one;
- Multiple Position Check – This variable reports when, in addition to the previously described visual sequence Visual Position Check, pilots carried out a check also on electronic equipment. The position equipment considered as additional check could have been, for example, the radar screen or the electronic chart plotter;
- Visual Direction Check – This variable reports the frequency at which the pilots gaze was moved from an external object within 30 degrees centred on the bow and another object in the same sector and / or a heading instrument (gyro repeater);
- Multiple Rotation Check – This variable reports the frequency at which pilots were monitoring the ROT of the vessel (Rate of Turn is the vessel yawing). It was considered a complete sequence when the gaze was shifted alternatively from an external object maximum 30 degrees off the bow and then directed back on the bow, combined with a visual check on the ROT sensor;
- Visual Speed Check – This variable reports the frequency at which pilots gaze was alternatively fixed on an external object at the beam of the vessel and then directed on a speed sensor (Log). It was also considered in the sequence if a reading from a speed indicator was carried out;
- Plan Check – This variable reports the frequency at which pilots were looking at a plan, previously completed, explaining how they would have accomplished the manoeuvre;
- Pilot Orders – This variable reports the frequency at which pilots gave an order, either to the Bridge personnel or, via radio, to the tugs.

## 3.4. Data collection

This paragraph describes how the data was collected during different stages of the research.

### 3.4.1. Stage 1 – Preparation

After an initial contact, generally by email, with those pilots interested to participate, a suitable appointment was arranged by the researcher to meet them individually. This meeting gave the opportunity to complete the necessary forms and to clarify any aspect of the research that, at that point, might have not been clear to the participant. If one session was not enough, others would have followed as necessary, until all the data required before the simulations was collected.

***Interview – Part 1***

The general Interview (see Appendix 10) was an initial video recorded face to face interview aiming at collecting some general information about the participant. Some of the questions were:

“Previous experience with similar ships to those used in the manoeuvres(if any)”;

“What is your relevant experience as a Pilot (briefly describe the type of ships and Ports you have been working and their peculiar characteristics in terms of manoeuvrability, limitations...)”;

Those general questions were asked during this initial meeting.

Some other questions, such as:

“What is you fatigue state today (briefly describe anything of interest happened in the last 24h that could influence the state of fatigue)”

“Caffeine Nicotine or other substances assumed recently (if any)”;

were specifically included to be answered on the day of the data collection in the simulator.

At the end of the General Interview, the next step was to complete the passage plans for the manoeuvres chosen for the research. The four manoeuvres, as described in section 3.3.2. were the same for all the pilots.

***The Detailed Manoeuvre Plan***

Part of the data collection carried out during the first meeting was the completion of detailed manoeuvring plans (DMP), one for each of the four manoeuvres. The DMP was an extensive explanation regarding how pilots thought would have performed the manoeuvre in the Simulator. In order to create and to obtain the record of such explanation in a numerical form a DMP table was compiled by each participant for each manoeuvre, prior to performing this manoeuvre on the Simulator. An example of a DMP compiled for a specific manoeuvre is provided in Appendix 6. This table can be seen as a more detailed version of the routine passage plan that normally is discussed between Pilot and Captain (including all the Bridge Team) before a ships enters into a port to be moored at a specific berth (Wild & Constable, 2013). The compilation of the mentioned table was obtained through a face to face exchange between the pilot and the researcher. The initial material provided by the researcher to the pilots includes also a facsimile of port navigational charts at the appropriate scale for each manoeuvre. An example of those charts is provided in Appendix 5. On these charts the initial and the final positions of the ship were indicated, specifying which side of the ship was required to be alongside at the end of the manoeuvre. At the beginning of the face to face exchange between the researcher and the pilot, only the chart and a brief explanation about the manoeuvre were provided (initial position, position of the berth, side to go alongside to). Pilots were invited to ask all the questions they deemed necessary to complete the planning. All the questions were collected in order to evaluate the elements considered with reference to the manoeuvres. There were no limitations whatsoever to the amount, specificity or topic of the questions. All the pilots received only the answers relative to their questions.

A list of all the questions asked is provided in Appendix 4. Table 11 shows the total amount of questions asked (and answered), grouped by main topics, as detailed in Appendix 4.

Table 11. Total number of questions asked by pilots while completing the DMP.

| **Questions Topics** | **Number asked** |
|---|---|
| Communications | 8 |
| Port Characteristics | 71 |
| Port Regulations | 31 |
| Port Services | 42 |
| Ship Crew | 13 |
| Ship Documents | 6 |
| Traffic | 17 |
| Weather and Environment | 167 |
| Ship Characteristics | 340 |
| Ship Bridge Equipment | 18 |

Pilots were allowed to require additional information at any time during Stage 1 until, in their opinion, the information was enough to allow them to proficiently complete the planning task summarised in the DMP.

Integral part of the DMP was also the drawing of nautical charts. Pilots were required to sketch ship's movements using those charts, identifying any associated element of interest with some precision. More specifically pilots were asked to complete their intended Detailed Manoeuvre Plan showing on the charts the sequential positions of the vessel (using waypoints). For each waypoint depicted on the chart, in the table, it had to be specified the speed profile, the use of ship's propulsion (main propulsion and thrusters), and external forces (tugs). Pilots had also to include a brief description of the effect of the environmental and hydrodynamic forces acting on the vessels (wind, current, tide, bottom and bank effects, etc..).

All the plans were collected before all the manoeuvres took place in the simulator. Prepared prior to the simulations these plans formed a comparative basis that was used to assess outcomes generated in the simulator. A full mission bridge simulator can record all the previously mentioned measurements (and more) with a high degree of accuracy at several samples per second.

### 3.4.2. Stage 2 – Execution in the Simulator

***Day at the Simulator***

Table 12. Sequence of events at the simulator.

| Phase Description | Recording Method |
|---|---|
| Introduction to the Facility<br>Explanation of the Daily Schedule | |
| Interview | Video |
| Equipment setup | |
| Baseline Recording - Before | EEG ECG |
| Exercise .....n | Video EEG ECG ETR Sim |
| Baseline Recording - After | EEG ECG |
| Nasa TLX and Forms Completion | Form |
| Debriefing | Video |

Once the initial data collection was completed and a DMP for each manoeuvre was provided, the participants were invited to spend a day at the Smartship Simulator to perform all four manoeuvres. The day before, an email (see Appendix 11) was sent in order to remind the participant of a few final details, especially concerning the use of caffeine and nicotine.

***Interview – Part 2***

As summarised in Table 12, the "simulator day" normally started with an introduction to the facility and an explanation of what would have been the daily schedule. Then the general interview (see Appendix 10) was completed, going through those questions that were specifically relevant to the simulator day. An example of one of those questions could be: "Caffeine Nicotine or other substances assumed recently (if any)".

***Equipment and Questionnaires***

Before starting the manoeuvres, pilots were fitted with the equipment necessary to record the physiological variables. Once the electro cardiogram, electro encephalogram and eye tracker equipment were tested, pilots were required to perform a very simple mooring manoeuvre with a vessel different from those used in the experimental manoeuvres. This first manoeuvre was used as a test and a familiarization manoeuvre in order to double check recording devices and have the pilots acquainted with the bridge environment and the navigation equipment. After this familiarization manoeuvre, the remaining manoeuvres that were previously planned in Stage 1, were used in random order as “hot manoeuvres”, so all the data of interest could be recorded. The recording of physiological variables was achieved in the least obtrusive possible way, using wireless and portable devices.

### ECG

For the recording of the Electrocardiogram signal, a Smartex® Wearable Wellness System® was used (www.smartex.it). This system, using two electrodes inbuilt in the fabric of the wearable t-shirt, was able to collect a full electrocardiogram. The system provided an input sensitivity of ±5 mV. The sampling rate was 250 samples per second, covering a bandwidth from 0.05 Hz to 30 Hz with a 12 bits resolution. Further analysis of the raw ECG signals, as better explained in section 3.5.3, was able to provide also the Heart Rate Variability (HRV).

Figure 11. Smartex ECG recorder - Wearable Wellness System (WWS)


### EEG

The electro encephalograph signals were obtained using an Emotiv® Epoc® wireless device (www.emotiv.com), featuring 14 channels: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF42. The sampling rate was 128 samples per second with a resolution of 14 bits (1 LSB = 0.51μV), covering a bandwidth from 0.2 Hz to 43 Hz, with digital notch filters at 50 Hz and 60 Hz.

Figure 12. Emotiv Epoc EEG recorder


### Eye Tracking, Audio Recording and Pupil dilation

Eye tracking goggles ASL® Mobile Eye XG® (www.asleyetracking.com) were used to record eye movements, audio and pupil dilation. The Eye Tracker goggles had a front HD camera that recorded what was in front of the subject at a rate of 30 frames per second. A second infrared camera was pointed towards the subject right eye and recorded the pupil movement and dilation (diameter measured in number of pixels). This system was then able merge the two channels, providing the gaze position through a red cross overlaid on the video recorded by the front camera. A calibration procedure, which determined where gaze was located within the scene, needed to be performed at the beginning of each recording. The audio was recorded by an inbuilt microphone.

Figure 13. ASL Mobile Eye XG eye tracking goggles

The audio allowed to record pilots orders and “thinking aloud” reports in the form of intentions shared with the rest of the Bridge. Audio recordings provided also the measurement of experienced workload based on the ordinal ratings of the self assessment Likert scale described in the next section.

### Self Assessment Likert Scale

Before the manoeuvres were started pilots were instructed about the use of a self assessment Likert scale which they had to refer to, to verbally report their level of “involvement or workload” during the manoeuvre. The scale, provided in Appendix 8 (and always kept in pilots’ sight), reported 7 different levels of “exercise difficulty”, meaning the personal level of workload experienced or effort necessary, in order to be able to manage the situation at the time of the question. The pilot, during the execution of the manoeuvre, was briefly reminded (every two minutes circa), with a quick question asked by the researcher (i.e. “How do you feel?”), to simply report a number according to the scale described below:

Level 7 was indicated as the level where the situation was felt so demanding, that was just about to be out of hand; level 6 was a challenging situation that required the complete attention of the pilot, working at almost 100% of his capabilities; level 5 was a situation requiring more attention than normal, but not felt as critical as level 6; level 4 wanted to depict a normal level of involvement where the pilot could feel perfectly capable to achieve the desired outcome with a necessary but comfortable level of effort (routine operation); level 3 was an easy condition offering no specific challenge, with required effort below the average; level 2 was a very comfortable, almost effortless situation; level 1 indicated a situation of “complete boredom”, with very little or no involvement.

Self assessment scales present some disadvantages: they can tend to be situation specific and may fail to take into account the individual’s learning, experience, natural ability and changes in emotional state, they also might reveal little in terms of the brain mechanism involved in task performance (Baldwin, 2003). Nevertheless, self assessment scales are relatively easy to administer and interpret and they do not require extensive training or equipment (Luximon & Goonetilleke, 2001). In light of all those considerations, an additional Mental Workload measurement was adopted: the NASA TLX.

#### NASA TLX

The National Aeronautics and Space Administration-Task Load Index (NASA-TLX). NASA-TLX (Hart & Staveland, 1988) is a multidimensional scale for which the overall mental workload is a function of 6 subscales: 1. Mental Demand (MI), 2. Physical Demand (PD), 3. Temporal Demand (TD), 4. Own Performance (OP), 5. Effort (EF), 6. Frustration Level (FR). Refer to section 2.3.2 of this thesis for more details. At the end of each manoeuvre, after the post exercise baseline recording, the pilots completed a NASA TLX questionnaire in a Microsoft® Excel® electronic format.

#### Simulator

The simulator was principally used to gather data related to performance. The simulator was capable to record a conspicuous amount of technical measurements at a rate of 5 Hz. Some of those measurements were for example: Piloted Ships data such as latitude and longitude, lateral and longitudinal speeds, rate of turn and heading, propulsion settings etc.., environmental conditions such as wind intensity and direction, current, etc..

***Experimental Manoeuvres***

Before the beginning of each experimental manoeuvre, the Pilot was informed about which of the four planned manoeuvres was about to come next. The relevant Detailed Manoeuvring Plan (compiled in Stage 1) was reviewed with the researcher in order to evaluate any possible doubt or additional question. A video recording, where the pilot went through the plan with the researcher, was finally captured. Before the simulation was started an initial physiological baseline recording of at least 5 minutes was carried out. After the physiological baseline recording, the simulation was started, allowing the pilot to execute the manoeuvre. During the manoeuvre there were no interruptions or suggestions by the researcher who was generally acting as the ship’s Captain or, when required by the specific context, as the member of the bridge team more suitable to interact with the pilot at that time. Immediately on completion of the manoeuvre, once the simulation was stopped, another physiological baseline of at least 5 minutes was recorded.

## 3.5. Data analysis

This section explains the initial synchronization procedures, the data filtering and the quantitative techniques that were used to conduct the data analyses.

### 3.5.1. Data files and formats

In Table 13 is summarised the list of raw data files that were collected after each manoeuvre. The collection of these raw data files, in different formats and independently time stamped by the clock of each single device, was the start of the data processing procedure, described in the following paragraphs.

Table 13. Measurements and output data formats

| Measurement | Device / Questionnaire | Nr of Files | Original Data file extension | Direct conversion in |
|---|---|---|---|---|
| ECG | Smartex WWS | 1 | .wws (proprietary) | .edf |
| EEG | Emotiv Epoc | 1 | .csv | .edf |
| Ship Data | Simulator | 4 | .csv | |
| Workload | NASA TLX | 1 | .xlsx | |
| Workload | ASL Mobile Eye XG | 1 | .avi (audio) | |
| Pupil Dilation | ASL Mobile Eye XG | 1 | .csv | .xlsx |
| Eye Tracking | ASL Mobile Eye XG | 1 | .avi (video) | |

### 3.5.2. ECG and EEG merging

The Smartex® Wearable Wellness System® was able both to stream the recorded signals via a Bluetooth connection and store them on a local SD card. At the end of each manoeuvre it was possible to download the recording in a proprietary format (.wws) from the portable device connected to the t-shirt. The software provided with the device, named SmartScope, was then able to export that file in an .edf format.

The Emotiv® Epoc® wireless device was only able to stream the recorded signals via a Bluetooth connection. The dedicated software, Emotiv Testbench, was then able to record a .csv file (comma separated values) using the streamed data. The .csv file was then converted in an .edf file using the open software EDFbrowser.

The European Data Format (EDF) is a standard file format designed for exchange and storage of medical time series. Being an open and non-proprietary format, .edf files are commonly used to archive, exchange and analyse data from commercial devices in a format that is independent of the acquisition system.

The Smartex portable recorder time stamped the ECG signal using its own internal clock, as well as did the Emotiv® Epoc® wireless device. This required to follow a synchronising procedure at different times during each manoeuvre, to be able to clearly identify those instants not only in the ECG and EEG recordings, but also across all the devices used in the research (see Appendix 7, points 3, 8, 10, 12, 14 and 17).

In order to clearly identify those instants, both on the ECG and the ECG signals, participants were required to tap their chest for few seconds and then, stronger, for three times, in order to leave on the signals easily recognisable marks.

Those marks allowed to synchronise and merge in the same EDF file, the ECG and EEG signals coming from the two different recording devices. Once the ECG and the EEG signals were synchronised on the same time line, showing consistent time stamp values, they could be processed separately to obtain the desired results.

### 3.5.3. ECG analysis

The ECG signal was processed using the free software Kubios HRV . Kubios software was used to visually inspect the raw ECG recordings and clean artefacts using the provided filtering functions. Identifying the instant of each heart beat on the time line, allows to obtain two continuous functions (one the inverse of the other) in the domain of time:

- The heart rate function, which reports the number of beats per minute at any instant in time;
- The Inter beat Interval (IBI) function, which reports the period (in milliseconds) recorded between each two consecutive heart beats, at any instant in time.

### *Heart Rate*

A continuous function reporting the heart rate (as beats per minute) in the domain of time was obtained for each manoeuvre performed in the simulator. Each function was divided into 5 sections or phases, as depicted in Table 8. Phase 0 and 5 were recorded before and after the execution of the manoeuvres, phases 1 to 3 were recorded during the execution of the manoeuvres.

### *Heart Rate Variability*

The IBI signal was the starting point for the Heart Rate Variability (HRV) analysis. As described in section 2.3.3, HRV is known as a non invasive technique to measure cardiovascular autonomic regulation (Hansson & Jönsson, 2006). It expresses the balance of the regulation of the sympathetic and parasympathetic nervous systems. HRV has been extensively exploited to study the association between psychological processes and physiological reactions (Berntson, 1997). The LF/HF ratio is an important parameter in the study of the power spectral density (PSD) of the inter beat intervals function (IBI). The IBI function has to be transformed from the domain of time into the domain of frequency using a Fourier Transform. Low Frequencies (LF – from 0.04 to 0.15 Hertz) and in the High Frequencies (HF – from 0.15 to 0.40 Hertz) are the adopted in the calculation of the LF/HF index (Malik et al., 1996). LF power component is connected with the sympathetic activities of the nervous system while the HF power component is more connected with the parasympathetic activities (Lord et al., 2001). Elevated values in the LF are associated with high stress (Van Amelsvoort, Schouten, Maan, Swenne, & Kok, 2000), resulting in higher scores in the LF/HF ratio. The strong correlation between Heart Rate Variability and Stress has been extensively documented in the literature (Thayer, Åhs, Fredrikson, Sollers Iii, & Wager, 2012). For the purposes of this paper, “stress” is defined as the transition from a calm state into an excited state, through the activation of the sympathetic system (Selye, 1980), considering as “stressors”, excessive intellectual, emotional and perceptual stimuli (Skinner & Simpson, 2002). The measurements of LF/HF ratio and the Heart Rate were specifically chosen, due to their sensitivity to work related stressors (Ritvanen, Louhevaara, Helin, Väisänen, & Hänninen, 2006; Van Amelsvoort et al., 2000).

Figure 14. ECG data processing to obtain LF/HF index

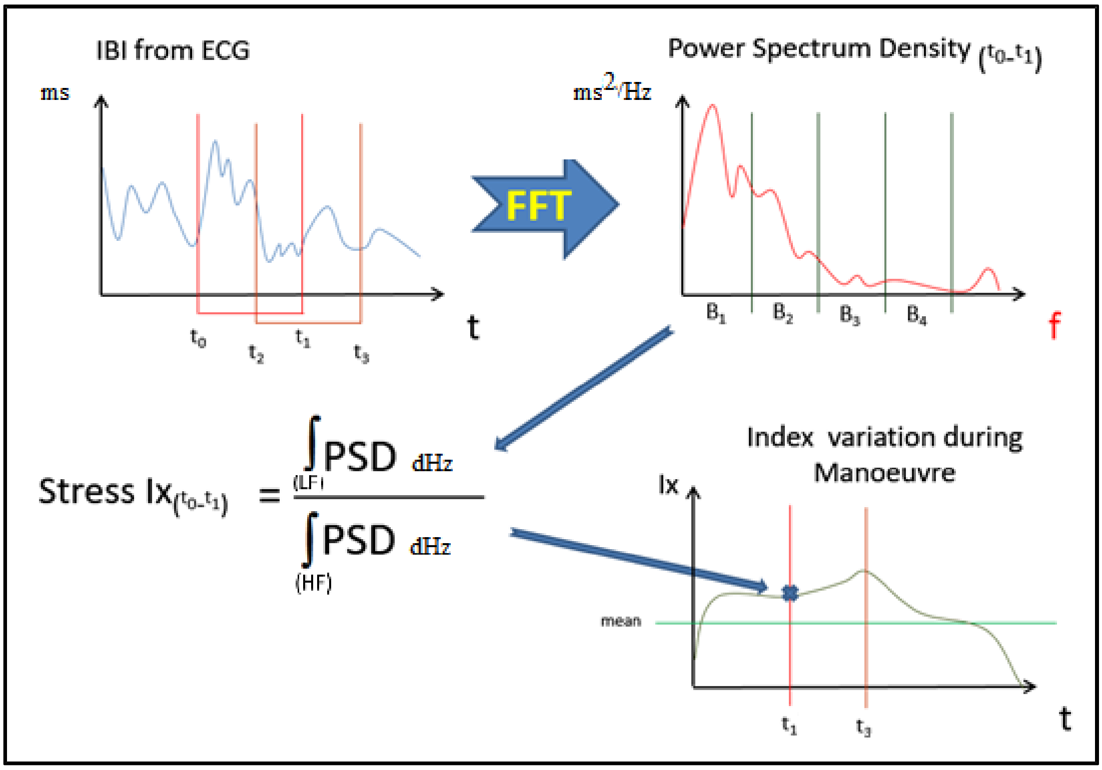



Figure 14 reports the process through which it is possible to obtain the LF/HF index. In the top left of the figure there is a representation of the IBI function in the domain of time (expressed in milliseconds), as obtained from the ECG signal. Initially a specific time interval included between the instants $t_0$ and $t_1$. can be considered. Using different mathematical techniques or algorithms, such as the Fast Fourier Transform, it is possible to obtain the power spectral density (PSD) function of the IBI signal, as recorded in the chosen interval ($t_0 - t_1$). The PSD function is a function in the domain of frequency and highlights the spectral components of the original IBI signal captured in the domain of time within the interval ($t_0 - t_1$). Obtaining the amount of "power" of the original signal in the desired bands, equals to calculate the integral of PSD function in those band intervals (expressed in Hz).

The index LF/HF at the instant $t_1$ is obtained as the ratio between the scalar obtained calculating the integral function of the PSD in the low frequencies (LF band – from 0.04 to 0.15 Hertz) over the scalar obtained calculating the integral function of the PSD in the high frequencies (HF band – from 0.15 to 0.40 Hertz). Choosing sequential intervals (also partially overlapping) of the IBI function, the LF/HF function can be plotted at different instants in time ($t_1$, $t_3$… $t_n$). In other words, to obtain a continuous function reporting the LF/HF values in the domain of time, we need to reiterate the calculations done for the interval ($t_0 – t_1$) on sequential intervals ($t_2 – t_3$, $t_4 – t_5$, $t_6 – t_7$, …, $t_{n-1}$ - $t_n$), covering the whole duration of the chosen IBI function.

In this research, the spectral analysis (in the domain of frequency) was obtained using the algorithms provided in the open source software HRVAS (Ramshur, 2010). HRVAS software was a graphical interface developed to implement functions for ECG analysis, built in Matlab scripting code (MathWorks, 2013). Those Matlab functions were extracted and re-integrated in newly developed scripts able to batch process all the ECG recordings collected during the manoeuvres.

An ectopic detection algorithm was used at the beginning of the analysis to exclude those values in the IBI functions that were exceeding 20% of previous sample value, or exceeding 3 times the standard deviation of the signal, or using a median filter where the tau value was set to 4. Those values were substituted using a “spline” interpolation between adjacent values. The cleaned IBI signals were de-trended, meaning that their average was reduced to 0. The spectral analysis of a signal, in fact, focuses on the variation of that signal in time, not on its absolute values.

For the spectral analysis (to calculate the LF/HF values) it was chosen a shifting window of 128 seconds. To obtain the continuous LF/HF function in time, all the calculations were reiterated shifting the 128 seconds window, of 1 second at a time. This process implied that 2 consecutive windows would have overlapped for 127 seconds. The spectral analysis of the 128 seconds IBI signal intervals was obtained using the using Welch's algorithm (Welch, 1967).

Reiterating the spectral analysis for each sequential 128 seconds interval, allowed to obtain, for each IBI recording, the LF/HF function in the domain of time. Similarly to what obtained with the heart rate functions, each LF/HF function could be divided into 5 sections or phases, as reported in Table 8. Phase 0 and 5 were recorded before and after the execution of the manoeuvres. Phases 1 to 3 were recorded during the execution of the manoeuvres.

### 3.5.4. EEG Analysis

The EEG signal is a representation of the brain electrical activity recorded by electrodes placed on the scalp. It has been used to assess operators workload for many years in both laboratory (Berka et al., 2007; Gundel & Wilson, 1992; Lei et al., 2009; Stevens, Galloway, & Berka, 2007) and applied settings (Kohlmorgen et al., 2007; G. F. Wilson, 2002). The EEG signals, recorded in the domain of time as electric potentials measured in microvolts ($\mu$V), can be analysed in the domain of frequency. Frequency bands for the study of EEG are generally defined as follows: Delta (below 4 Hz), Theta (4 - 7 Hz), Alpha (8 – 13 Hz), Beta (13 – 36 Hz). For the purposes of this research, the bands Beta 1 and Beta 2 were considered.

The spectral analysis in the chosen bands (Beta 1 and Beta 2) was performed following a similar process to what adopted for the HRV analysis (see section 3.5.3). One of the major differences was that for the EEG, 14 different signals coming from 14 different electrodes, had to be analysed at the same time.

The signals were broadcasted via Bluetooth, from the headset to a computer able to record them. The 14 signals were checked for integrity, meaning that if few samples were missing for bad reception, they were replaced using a linear interpolation between the values at the extremities of the gap. Signals were cleaned from outliers, artefacts and spikes, using a dedicated pre processing filtering function able to detect differences in values between consecutive samples: exceeding 20% of previous sample or exceeding 3 times the standard deviation of the signal or using a median filter where the “tau” value was set to 4.

Once the signals were cleaned from artefacts, they were also de-trended so to be ready for the following spectral analysis. Similarly to what accomplished in the processing of the HRV, it was chosen a shifting window of 128 seconds. To obtain continuous values of the Bands Beta 1 and 2 in the domain of time, all the calculations were reiterated shifting the 128 seconds window, of 1 second at a time. This process implied that 2 consecutive windows would have overlapped for 127 seconds. The spectral analysis in the bands Beta 1 and Beta 2 of the 128 seconds intervals extracted from the EEG signals (on 14 different channels) was obtained using Welch's algorithm (Welch, 1967). The Beta 1 values for all 14 channels were obtained as a fraction of 1. Such fraction expressed the spectral power measured in the band Beta 1 (calculated integrating the PSD function in the Beta 1 interval), divided by the total power measured in all the other bands considered (Delta + Theta + Alpha + Beta). The same calculation was performed on all the 14 channels for the band Beta 2.

Reiterating the spectral analysis for each sequential 128 seconds interval, allowed to obtain, for each channel, the Beta 1 and Beta 2 functions in the domain of time. In order to obtain a single Beta 1 function in the domain of time for each manoeuvre, it was considered the average of all the Beta 1 scalars calculated on the 14 different channels, for every shifting window. The same process was followed for the Beta 2 function. The Beta 1 and Beta 2 average functions, so obtained, could be then considered into the 5 sections or phases into which each manoeuvre was divided, as reported in Table 8. Phase 0 and 5 were recorded before and after the execution of manoeuvres, phases 1 to 3 were recorded during the execution of the manoeuvres.

### 3.5.5. Pupil Dilation

The pupil dilation was recorded in a .csv file, reporting the pupil diameter measured in number of pixels and timestamped consistently with the video recording carried out using the eye trackers. The diameter was collected by the system 30 times per second. When data was lost for temporary interruption of the pupil tracking, it was replaced using a linear interpolation between the values at the extremities of the gap. Signals were cleaned from outliers, artefacts and spikes, using a dedicated a pre processing filtering function able to detect differences in values between consecutive samples: exceeding 20% of previous sample value or exceeding 3 times the standard deviation of the signal or using a median filter where the “tau” value was set to 4.

The pupil dilation scores were directly collected as a function in the domain of time for all the manoeuvres. For each manoeuvre only 3 sections or phases were considered (1, 2 and 3), as reported in Table 8. Phase 0 and 5 were not recorded before and after the execution of manoeuvres, since the eye trackers were not active at that point in time.

### 3.5.6. Quantile Normalisation of physiological scores

In order to be able to compare physiological data, relative to different measurements and several subjects, a quantile normalization of the scores was adopted (B. M. Bolstad, Irizarry, Åstrand, & Speed, 2003). The normal distribution was selected as reference distribution for the transformation. This procedure was applied to EEG (Heart Rate and LF/HF), EEG (Bands Beta 1 and 2) and Pupil Dilation.

The empirical cumulative distribution of the raw scores was calculated in Matlab (MathWorks, 2013). All the scores (of each one of the dependent variables) recorded in the four manoeuvres performed, were considered all together for each subject. For each raw score (considering the same subject and all the four manoeuvres together) it was then possible to calculate its percentile value. Using the percentile value “X” it was possible to obtain the correspondent value “Y” of a normal distribution (mean = 0 and standard deviation = 1), using the inverse function of the empirical cumulative distribution. This process was reiterated for each subject and for each dependant physiological variable. This process allowed to obtain a normal distribution of all the scores recorded by any subject, for any specific physiological variable and in all the 4 four manoeuvres. The subjects’ normal scores could then be used to describe a continuous function in the domain of time that could be used for further analysis. Averages across participants, on the three considered factors in the study, were then calculated using previously normalized scores.

### 3.5.7. Eye Tracking and Audio Recording

Eye movements and audio were recorded using a pair of light weight eye tracking goggles (ASL® Mobile Eye XG®). The eye-tracking video and audio recordings obtained from each manoeuvre, were coded using the open source video annotation software Anvil (Kipp, 2010). A set of labels, able to identify specific elements that pilots directed their gaze towards was defined, creating a list of behavioural markers (BM). The complete list of behavioural markers adopted is provided in Appendix 12. Due to the long time necessary to complete the manual coding for the whole duration of the manoeuvres, a sampling strategy was applied in order to obtain the video coding from specific parts of the manoeuvre. A total duration of 20 minutes of video coding was obtained for each manoeuvre, considering 4 different sections of 5 continuous minutes. The section's location in each manoeuvre was guided by the following criteria:

Table 14. Locations of video coded sections in each manoeuvre

| **Approach** | | | **Swing** | | | **Closing** | | |
|---|---|---|---|---|---|---|---|---|
| Coding Window 1 | GAP | Coding Window 2 | GAP | Coding Window 3 | GAP | GAP | Coding Window 4 | GAP |

Each coded section had always a minimum duration of five minutes (unless the whole duration of that specific manoeuvring phase was less). In the swing and in the closing section the video coding was always placed in the middle of the identified window, while for the Approach section, it was placed at the very beginning and at the end, finishing with the limit of the swing section. Gaps, of course, had to vary accordingly, depending on the duration of the different manoeuvre sections. The distribution of coding windows in the manoeuvre is depicted in Table 14.

For each window, a complete coding of the gaze location was achieved. Data files so obtained simply reported a list of coding labels, indicating what and/or where (as relative bearing from the bow of the ship) the gaze was aiming. Each coding label was timestamped with a starting and ending instant. An example of a fragment of one of those files is provided in Appendix 13. Defining and searching a specific sequence of behavioural markers, meant to scan those files and identify if a pre-defined list of labels appeared, one after the other according to a chosen order, within a certain amount of time.

In this research certain sequences, obtained combining different behavioural markers, were specifically chosen as meaningful, based on researcher experience as shiphandler. These sequences are reported in Appendix 9. In Table 15 it is explained the meaning, from a shiphandling point of view, of each one of those sequences or dependant behavioural variables defined in section 3.3.3 (Behavioural Variables).

Table 15. Behavioural Markers and Targeted Sequences

| BM / Sequence Name | Shiphandling objective |
|---|---|
| Visual Position Check | Looking at different objects which have a relative angular distance around 90° allows to gain spatial awareness about ship position. |
| Multiple Position Check | Spatial awareness gained through the previous sequence of BM can be reinforced through the use of positioning equipment (Radar screen, chart plotter, Pilot Portable Unit) |
| Visual Direction Check | Directional awareness can be gained gazing at the foremast and at objects within 30° off the bow. |
| Multiple Rotation Check | The rotation of the ship can be checked also adding to the previous sequence a glance on a rate of turn sensor (ROT indicator) |
| Visual Speed Check | A glance at the beam of the vessel, followed or anticipated by a glance on a speed sensor (LOG or GPS SOG indicator) is used to appreciate the speed of the vessel, through the perception of the relative motions of lateral objects |
| Plan Check | Glance at the plan, is generally given when the pilots require some additional information not immediately available to his knowledge. |
| Pilot Orders | Pilot uses orders to direct the bridge team or tugs in assistance (via radio) |

A script was developed (Excel macro), able to batch scan all the video coding data files and obtain, for each manoeuvre, at intervals of 15 seconds, what was the number of the targeted sequences or behavioural markers (see Table 15) in the previous 60 seconds. In this way, every dependant behavioural variable could be graphed as a continuous function in the domain of time in each coding window.

### 3.5.8. Self Assessment Likert Scale

Exploiting the same video coding software (Anvil) it was also possible to code the audio recordings. This was done for the entire duration of each manoeuvre in order to keep track of the answers provided by the pilots when asked to report their perceived level of involvement or workload experienced, on a Likert scale from 1 to 7 (Appendix 8). Connecting those ordinal evaluations (regularly provided every couple of minutes), it was possible to graph for each manoeuvre a continuous function in the domain of time of the workload level experienced and reported by pilots.

### 3.5.9. NASA TLX

At the end of each manoeuvre, after having finished the “Post Base Line” physiological data recording, every pilot was required to complete a NASA TLX questionnaire relative to that manoeuvre. The questionnaire was administered to the pilots in the form of an Excel workbook. A sliding ruler (with 21 possible positions) was assigned to each of the six scales to record the raw scores. The initially required weighing procedure of the scales was performed only once and at the end of the first experimental manoeuvre. Using the results obtained from the initial weighting procedure and combining them with the raw scores recorded by the sliding rulers, it was possible to obtain, for each manoeuvre, the weighted scores in the 6 sub scales and the total NASA TLX score. These results, referring to the whole manoeuvre, could not be used for analysis on the factor “phase”.

### 3.5.10. Simulator Recordings

At the end of each manoeuvre, the Smartship simulator was able to provide .csv time stamped files reporting data relative to the ships used, weather conditions, forces involved etc. All the data was recorded in files at intervals of 0.2 seconds (5 Hz). For the purposes of this research it was taken into consideration specifically the data relative to the piloted ship (own ship) and the tugs used (when made available in the exercises). Among the many parameters recorded, in Table 16 it is reported a list of those considered in this thesis and used for the calculations reported in paper II (see section 6.2):

Table 16. Raw Simulator Data

| | | |
|---|---|---|
| Ship | Heading | Deg True |
| | Rate of Turn | Deg/min |
| | Lateral Speed | Knots |
| | Longitudinal Speed | Knots |
| | Engine Setting (controls) | Fraction of 1 |
| | Engine Power | HP |
| | Bow Thruster Setting (controls) | Fraction of 1 |
| | Bow Thruster Power | HP |
| | Cross Track Distance | meters |
| Tugs | Tension on lines | tons |

The simulator raw data was used to be compared with the estimations provided by the pilots through the completion of their Detailed Manoeuvring Plan (one DMP for each manoeuvre) collected in Stage 1 of the research (see paragraph 3.4.1.). Dependant

Performance variables were obtained considering specific differences between the estimations (DMP) and the actual measurements recorded by the simulator during the execution of the manoeuvres (see section 3.3.3 – Performance Variables). The comparison between DMP and simulator data, was achieved synchronising the DMPs with the simulator replays. For each step or waypoint described in the DMP by the pilot, it was inserted the relevant simulator timestamp (obtained by inspecting the replays). The DMP values considered at any waypoint, were kept constant until the next waypoint, where new estimation values were considered. Any difference (used as performance variable) was then calculated as a continuous function of time. Those performance variables were: the distance from the intended track and the effective ship track (Cross Track Distance); the difference between the intended speed over the ground and the recorded ship speed; the difference between the intended use of power and the recorded main engine and thrusters power, the difference between the expected use of tug force and their actual use.

Results, obtained as the difference between two functions (the estimation and the execution), were graphed as a function in the domain of time in paper II.

### 3.5.11. Missing Data

During the execution of the 40 manoeuvres, a total of four crashes were recorded. A “crash” was an impact or a grounding that required the interruption of the simulation and data collection. All the crashes were experienced during the manoeuvres at the higher level of difficulty, three during the manoeuvres in the foreign port and one in the homeport. Since, after a crash, the simulator had to be stopped and reset, no further data collection for that manoeuvre could be possible. More specifically, two crashes happened during the approach (first phase of the manoeuvre), so no data could be collected during the following two phases (swing and closing). Two crashes were recorded during the swing, so no data could be collected for the following (closing). Three impacts were also experienced (one in the foreign port with the easier level of difficulty and two during the swing in both ports at the higher of difficulty). An “impact” was classified as a contact of the vessel with another ship or port infrastructures that did not impede the continuation of the manoeuvre. In addition, due to a temporary malfunction of the EEG headset, three manoeuvres did not have an available EEG recording. The unavailable data was left missing in the dataset used for following analysis.

### 3.5.12. Synchronization of all dependant variables

The data processing so far conducted was able to provide the dependant variables considered in this research, for the whole duration of each manoeuvre, as continuous functions in the domain of time.

As explained in section 3.5.2. an initial synchronization was achieved between EEG and ECG on a common time line. The pupil dilation data was recorded by the eye trackers. The output .csv files reported a time stamp based on the eye tracker video recording clock. The coding of the video clips required to use a dedicated application (Anvil) which was time stamping the output data files using its own internal clock. The simulator related data was generated, referring to the simulator internal clock.

To be able to perform further analysis, it was necessary at this stage to synchronise all the considered dependant variables (originally timestamped using four different clocks) using a common timeline.

The simulator time was chosen as the reference timeline. All the other timestamps were shifted, using offsets obtained through time check procedures systematically carried out during the recordings (see Appendix 7, points 3, 8, 10, 12, 14 and 17).

Through the use of macro scripting, for each manoeuvre, all the dependent variables output files were merged into a single Excel data file, with time stamps synchronised and consistent with the simulator clock.

### 3.5.13. Analysis of Variance

To evaluate differences between the means, three independent variables “Difficulty”, “Port” and “Phase” were adopted as factors (see paragraph 3.3.2.) for the 3 Way ANOVA of all the depended variables so far described (see paragraph 3.3.3.).

From the analysis of the video recordings and the simulator replays, it was possible to identify at which instant in time each manoeuvre was transitioning between phases. To provide an example, it was used as the instant when the Swing started, when there was a clear order or indication given by the pilot that he was committing to the vessel rotation. This was also cross checked with a clear indication that the ROT was increasing and actions of tugs and ship propulsion were consistent with what was required to induce the swing (rudder angle to one side, tugs lifting off and / or pushing at the beam of the vessel, etc..). The start of the Closing was identified not only through pilots’ direct indications, but also cross checking when the ROT was minimised again (swing completed and vessel stabilised on a constant heading). Also other elements were considered, such as actions with ship propulsion and tugs, consistent with the end of the rotation and closing to the berth.

Clearly identifying for each manoeuvre, the instants when the Swing and the Closing phases started, allowed to section the functions of all the dependant variables (physiological, workload and performance) within the three manoeuvring phases. Each function of each dependent variable within each phase of the manoeuvre was averaged.

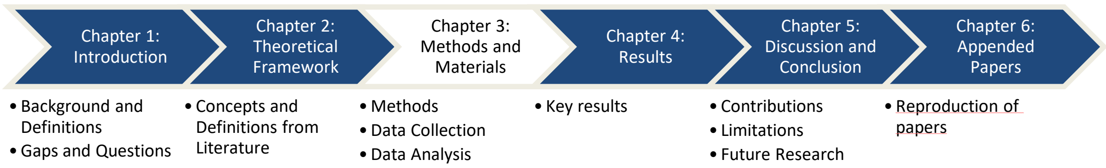
A scalar value for each dependant variable was obtained for each one of the three phases of the forty manoeuvres. Finding the time limits of the phases, allowed also to set the limits of the four coding windows that were adopted to conduct the analysis of the dependant behavioural variables (as detailed in paragraph 3.5.7). For each of those coding windows, the maximum value recorded in every dependent behavioural variable was chosen for the following analysis of variance.

Table 17. Papers and statistical analysis conducted

| Paper Title | Analysis Conducted | Independent Variables or Factors | Dependent Variables | Cases Number |
|---|---|---|---|---|
| A Comparison of Marine Pilots' Planning and Manoeuvring Skills: Uncovering Mental Models to Assess Shiphandling and Explore Expertise | ANOVA 3-WAY Between Subjects | Difficulty - 2 Levels<br>Port - 2 Levels<br>Phase - 3 Levels | XTD - Cross Track Distance<br>SpdEst - Estimation of Speed<br>EngEst - Estimation of Engine Use<br>ThrEst - Estimation of Thruster Use<br>TugEst - Estimation of Tugs Use | 10 Subjects X<br>4 Manoeuvres X<br>3 Phases<br>= 120 |
| Measuring mental workload and physiological reactions in marine pilots: building bridges towards redlines of performance. | ANOVA 2-WAY Between Subjects | Difficulty - 2 Levels<br>Port - 2 Levels | NASA TLX | 10 Subjects X<br>4 Manoeuvres<br>= 40 |
| | ANOVA 3-WAY Between Subjects | Difficulty - 2 Levels<br>Port - 2 Levels<br>Phase - 5 Levels | ECG - Heart Rate<br>ECG - LF/HF<br>EEG - Beta 1<br>EEG - Beta 2 | 10 Subjects X<br>4 Manoeuvres X<br>5 Phases<br>= 200 |
| | ANOVA 3-WAY Between Subjects | Difficulty - 2 Levels<br>Port - 2 Levels<br>Phase - 3 Levels | Likert Workload Scale<br><br>Pupil Dilation | 10 Subjects X<br>4 Manoeuvres X<br>3 Phases<br>= 120 |
| Interpreting changes in marine pilots' perceptual cycle through gaze detection patterns | GEE 2-WAY Repeated Measures | Difficulty - 2 Levels<br>Coding Windows - 4 Levels | Visual Position Check<br>Visual Direction Check<br>Visual Speed Check | 10 Subjects X<br>4 Manoeuvres X<br>4 Coding Windows<br>= 160 |
| Interpreting changes in marine pilots' perceptual cycle through gaze detection and speech patterns<br><br>- Unpublished - | ANOVA 3-WAY Repeated Measures | Difficulty - 2 Levels<br>Port - 2 Levels<br>Coding Windows - 4 Levels | Visual Position Check<br>Multiple Position Check<br>Visual Direction Check<br>Multiple Rotation Check<br>Visual Speed Check<br>Plan Check<br>Pilot Orders | 10 Subjects X<br>4 Manoeuvres X<br>4 Coding Windows<br>= 160 |

The results that will be described in details in chapter 4 (see Figure 15), were obtained performing an analysis of variance using datasets of different dimensions, as summarised in Table 17.

Figure 15. Progress Tracker – Moving to Chapter 4



# 4. Results

This chapter reports the findings obtained from the analysis detailed in the previous chapter. Table 18 summarises the key results of, and the messages conveyed in, the five papers reproduced in Chapter 6.

Table 18. Papers key results and conclusions

| Paper | Aim of the paper in relation to the project | Data collected | Key results and conclusions | Research questions |
|---|---|---|---|---|
| I | The Development of a Shiphandling Assessment Tool (SAT): A Methodology and an Integrated Approach to Assess Manoeuvring Expertise in a Full Mission Bridge Simulator | - Literature review | - Definition of Experimental Design<br>- Definition of research context<br>- Definition of Methodology<br>- Description of infrastructures, tools and equipment necessary. | 1, 3, 5 |
| II | A Comparison of Marine Pilots' Planning and Manoeuvring Skills: Uncovering Mental Models to Assess Shiphandling and Explore Expertise | - Literature review<br>- Pilot's plans (DMP)<br>- Simulator data | - Identification of critical areas in manoeuvres where pilot expectations departed from actual results<br>- Implications for shiphandling safety limits and margins | 1, 2 |
| III | Measuring mental workload and physiological reactions in marine pilots: building bridges towards redlines of performance. | - Literature review<br>- Physiological measurements<br>- Workload measurements | - Comparison between self assessment scales and physiological reactions<br>- Identification of most suitable and unobtrusive measures to indirectly monitor subjective workload<br>- Implications for shiphandling safety | 3, 4 |
| IV | Interpreting changes in marine pilots' perceptual cycle through gaze detection patterns. | - Literature review<br>- Eye tracking video recordings | - Identification and collection of behavioural markers able to highlight changes in shiphandling techniques<br>- Example of quantitative analysis of observable behaviours in maritime domain | 5, 6 |
| IV<br>AV | Interpreting changes in marine pilots' perceptual cycle through gaze detection and speech patterns | - Literature review<br>- Eye tracking video and audio recordings | - Identification and collection of behavioural markers able to highlight changes in shiphandling techniques<br>- Example of quantitative analysis of observable behaviours in maritime domain | 5, 6 |

Those key results will be presented in section 4.1. Section 4.2 will report a summary

## 4.1. Presentation of the key results

This section presents the findings as well as the main conclusions reported in each paper, and explains how they address the research questions formulated in Chapter 1.

### 4.1.1. Paper I: 'The Development of a Shiphandling Assessment Tool (SAT): A Methodology and an Integrated Approach to Assess Manoeuvring Expertise in a Full Mission Bridge Simulator'

**Paper I.** In this specific study there were no empirical results presented. The article described how the proposed methodology, aiming to objectively evaluate groups and / or individuals' shiphandling capabilities, would have been initially tested in a Full mission Bridge Simulator. Given the complexity and the broadness of the concept defined as "Shiphandling Expertise" (as presented in Chapter 2), several different objective measurements were isolated, to allow the researcher to empirically quantify different levels of Pilots' "expert performance". Those measurements took into account pilots' planning and forecasting capabilities through the compilation of a Detailed Manoeuvre Plan (DMP). The DMP was a more advanced version of the currently used "Pilot Master Exchange Briefing". The obvious difference with the latter was that the proposed DMP provided a more stringent and quantifiable way to report the information required, to allow a comparison between planned action and execution. As a results several performance variables could be proposed as concurring measurements able to evaluate different levels of technical shiphandling capabilities. Several and unobtrusive self reporting and physiological variables were also introduced, able to provide converging measurements of the workload experienced by pilots. Those measurements were thought to provide an insight of the pilots' level of involvement or difficulty subjectively experienced, helping to better understand end evaluate correlations with following performance outcomes. Those variables were compared with an overall subjective assessment of the manoeuvre, obtained from the administration of a NASA TLX form. The paper described also how behavioural variables, obtained from eye trackers audio and video recordings, would have been considered to measure changes in working strategies adopted. Basically, the paper provided a general description about how the authors intended to translate known psychological constructs into empirical measurements.

### 4.1.2. Paper II: 'A Comparison of Marine Pilots' Planning and Manoeuvring Skills: Uncovering Mental Models to Assess Shiphandling and Explore Expertise'

**Paper II**. Differently from the previous one, this paper was able to offer empirical results. This paper explained in details the analytical approach followed to compare pilots' detailed manoeuvre plans (DMP) with the executed manoeuvres. The DMP collected were considered as the tangible numerical translation of pilot's mental models. The expectation was that proficient pilots would have been able to provide plans that had a high degree of consistency with the execution. The aim of the study was to provide a clear example of an objective procedure able to analytically quantify the match between plans and execution. Performance variables were defined accordingly (see section 3.3.3.), and overall results confirmed the expectations: pilots were generally able to perform according to their plans, showing only a limited number of differences in the scores recorded in the different ports, at different levels of difficulty and during different phases of the manoeuvres. This result was also confirmed by Pearson correlation coefficients calculated between plans and execution. More specifically, correlation coefficients between manoeuvres with the same level of difficulty further showed consistency in the way those exercises were designed, hence approached and performed. The significant differences encountered, instead, pointed the attention to possible areas of improvement where pilots' approach to the manoeuvres could be discussed, reconsidered and modified.

The analysis of the cumulative distribution function, for example, showed how it was possible to define a distance in meters once a certain probability to remain within that distance from the intended track was chosen. Choosing for example to remain 80% of the times within a certain distance from the intended track, would require different distances in the different phases of the manoeuvres. It was calculated that during the swing that distance had to be at least 100 metres, while for the rest of the manoeuvre could be reduced to 50 metres. These estimations become critical when remaining within a certain distance from the intended track may make the difference between a safe movement or an incident. Analysis conducted on the speed showed how the use of a different type of propulsion decreases pilots' accuracy to estimate vessel's motion. The analysis conducted on the use of the main engine power was able to highlight how pilots expected to use the propulsion much less than experienced in the simulation. In the "swing" the difference between the planned and the effective use, reached values of 50%, when the engine was working already up to 80% of its maximum power. Pilots' plans did not consider to use the main engine that much, nor so close to its maximum availability. This observation suggested that the manoeuvres could have required a different approach in that particular phase to increase safety margins. These examples show how the methods and the analysis introduced in this paper can help to improve not only the understanding of shiphandling but can help shiphandlers to better identify critical manoeuvring practices, allowing the development of more effective and safer techniques.

### 4.1.3. Paper III: 'Measuring mental workload and physiological reactions in marine pilots: building bridges towards redlines of performance.'

**Paper III**. This paper reported results related to mental workload, a self reporting scale, a questionnaire and physiological responses. Several physiological variables were collected and analysed in order to obtain measurements that could be compared to scores from NASA TLX and a second self-assessment workload Likert scale. Results obtained from measuring ECG, EEG, and pupil dilation provided some indications that physiological variables correlated to scores obtained from self-assessment scales. Light correlations were highlighted specifically between the self-assessment Likert scale and the heart rate ($r = .334$) or between the self-assessment Likert scale and the pupil dilation ($r = .243$). In the study showed also how increasing difficulty induced a significant increment in pilots' physiological responses, particularly in the HR (23% of increment between easy and difficult manoeuvres). The part of the manoeuvre that elicited the strongest reaction was the swing. Controlling the safe rotation of a large vessel in constrained waters and in critical environmental conditions, challenged even expert pilots, and this was consistently shown not only in pilot's verbal reports, but also by all their physiological responses. The inclusion of a novel or unknown port in the research protocol did not show a significant effect on increasing the experienced workload. Marginal increases in self-assessed workload were not reflected in similar changes in the physiological data. This result may initially suggest that the use of pilots own port did not offer an advantage in the way manoeuvres were performed. It has to be reiterated, though, that four crashes happened during the conduction of the difficult manoeuvres (preventing the complete collection of scores in those manoeuvres) and three of these crashes occurred in the virtual port, affecting the mean of the scores collected during those manoeuvres (see paragraph 3.5.11.).

The study was able to demonstrate how different manoeuvring conditions were able to influence pilots workload and related physiological reactions. In summary, the results showed how the use a full mission bridge simulator could be suitable to create different scenarios inducing different levels of engagement or mental workload and physiological responses. Those responses could be used as unobtrusive, indirect measurements of mental workload. By investigating those responses, we may better define what are generally considered "easy" or "difficult" conditions. The aim would be to move us some of the way towards identifying an upper "redline" for the task demands of marine pilots, in the context of available resources.

### 4.1.4. Paper IV: ‘Interpreting changes in marine pilots’ perceptual cycle through gaze detection patterns.’

**Paper IV (Submitted) and Paper IV (Additional Version).** This study was included in this thesis in two different versions. The first version (reported in section 6.4) was the submitted paper, reporting a General Estimating Equation analysis, performed on the factors “Difficulty” and “Phase” for the dependent variables: position, direction and speed check. In this version the factor “Port” was considered as a repeated measure within subjects. The second version (inserted in section 6.5) was a previous review of the submitted paper. This unpublished version reported the results obtained from a 3 way ANOVA performed on the factors “Difficulty”, “Port” and “Phase for all the dependent variables listed in section 3.3.3 (Behavioural Variables). Publishing constrains ruled in favour of the simplified version (section 6.4). For the purpose of this thesis, though, this paragraph presents the key points obtained from the additional version, since considered more comprehensive than those reported in the submitted version.

In the study, a set of Behavioural Markers (BM) were defined (see Appendix 12) in order to code video clips obtained from eye trackers worn by participants. An example of video / audio coding output is provided in Appendix 13. Combining some of those behavioural markers in meaningful sequences, a list of dependent variables was created (see Appendix 9). The study reports how those behavioural variables showed significant results depending on manoeuvring conditions.

The variable “Visual Position Check” significantly increased from the very beginning of the manoeuvres until the end. This significant main effect was explained by the fact that, at the beginning, when the exercise was started, the ship was in a known position and making way, so for pilots it was more important to monitor the direction where the ship was going instead of checking the actual position. Later on in the manoeuvre knowledge of position became more relevant. For the swing the vessel had to be positioned exactly where it had enough clearance to rotate with sufficient distance from surrounding obstructions. Results showed that the frequency of Visual Position Check reached its peak specifically during the swing (coding window 3 - see paragraph 3.5.7. Table 14), significantly more for the difficult manoeuvres.

The variable “Multiple Position Check”, another dependent variable related to checking ship’s position using equipment, also showed a generalised and significant increase in trend. More importantly, manoeuvres in the homeport showed statistically lower scores at the difficult level compared to those in the foreign port. This result provided support to the idea (see section 2.1.1 – Domain Specificity and 2.1.9 – Local Knowledge) that familiarity with the homeport could enable pilots to rely more on their knowledge of features observed in the environment, rather than positioning equipment.

The variable "Visual Direction Check" represented a strategy through which pilots monitored vessel direction of motion. This variable significantly decreased from the very beginning of the manoeuvre until the "Swing" to increase once again in the closing. It was argued that it was observed almost an opposite trend in comparison with what was noted with the Visual Position Check. Our explanation was that at the beginning of the manoeuvres, the vessel was sailing at an appropriate speed. Pilots' concern was more to check that the direction was correct more than position, since an incorrect course would necessarily mean that the future position of the vessel will be incorrect.

The Variable "Multiple Rotation Check", related to the perception of vessel's rotation, had significant results on the factor Difficulty. In the easy manoeuvres more checks were performed than in the difficult ones. It was believed that this outcome was related to the different type of vessel propulsion available in the two levels of difficulty. In the easy manoeuvres the adopted vessel had a controllable pitch propeller (CPP). This propeller changes the angle of attack of its blades to change the thrust of water (or even revert it to go astern, backwards). Pilots had to reduce the propeller blade pitch to 0, in order to reduce the ship speed. This action, needed to reduce speed, reduced also the effectiveness of the rudder (shielded by the rotating propeller). Pilots had to closely monitor such hydrodynamic effect, in order to timely correct it. Results showed also how the checks on the ROT (rate of turn) indicator decreased throughout the whole duration of the manoeuvres, having their peak at the beginning.

The variable "Visual Speed Check" showed a general increase in frequency throughout all the manoeuvres. To safely moor a vessel alongside a berth, pilots had to progressively reduce the speed of the ship in order to arrive alongside with minimal momentum, calibrating the landing until the final touch on the fenders.

The variable "Plan Check" highlighted how the higher values were found at the very beginning of the manoeuvre, with the Closing (at end of the manoeuvre) significantly recording the lowest scores. The variable "Plan Check", through a significant main effect on the factor Port, clearly showed how the adoption of an unfamiliar port as experimental condition, forced pilots to refer more to the charts included in the DMP. Pilots referred more to those charts in the foreign port, to locate elements useful for ship positioning (transits, navigational aids, etc..) and to double check that they were following the intended plan.

The variable "Pilot Orders" showed in the difficult manoeuvres a significant effect having the highest scores just before and during the Swing. It is important to remember that the difficult manoeuvres were performed not only through giving orders to the bridge personnel but also to tugs. Pilots controlled the tugs giving them orders using a VHF radio. The use of tugs was particularly relevant when rotating and translating sideways a ship. It is not surprising then that the highest rate of orders was achieved just before and during the swing phase (coding windows 2 and 3).

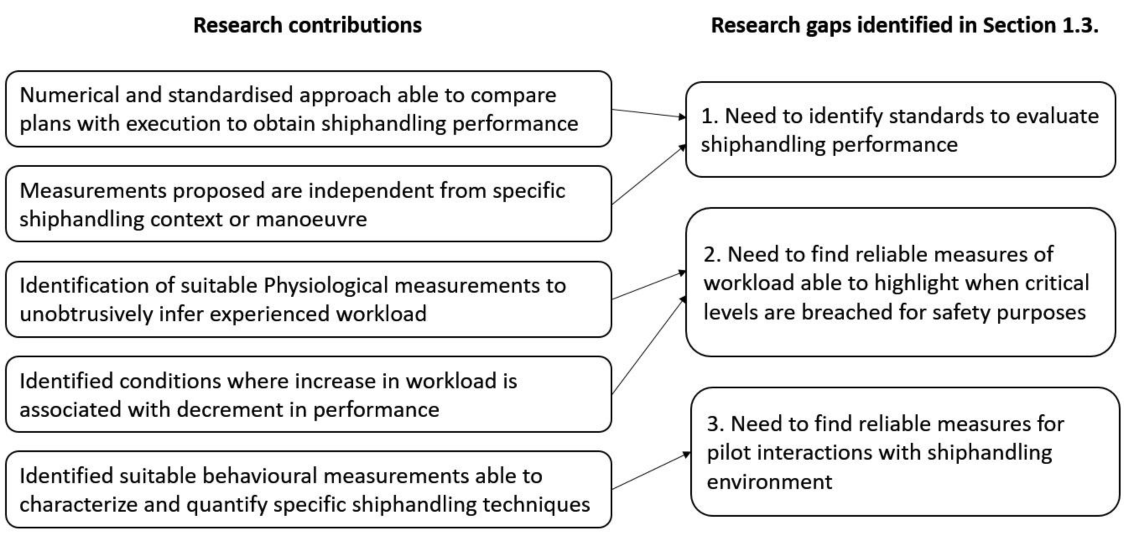
## 4.2. Summary of results

After having presented the research findings specific to each paper in Section 4.1, this section summarises the key results of those studies as a whole. In this section the obtained results were related with the three research gaps identified in Section 1.2.

1. The methodology demonstrated that it was possible to translate a manoeuvring plan into a numerical format. Mental models were compared to the execution of manoeuvres. This process was able to provide performance outcomes.
2. The methodology was not context specific, but could be applied to different scenarios.
3. Results have provided indications about feasibility to adopt specific physiological measures to unobtrusively monitor workload.
4. Specific experimental conditions elicited higher levels of workload and reduction in performance.
5. Behavioural measurements selected in this research, were able to capture elements of pilots' perceptual cycle.

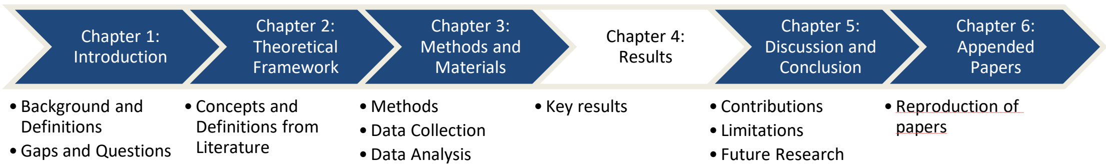
Figure 16 illustrates how the above mentioned results addressed the research gaps identified in Section 1.2.

Figure 16. Filling the research gaps



This chapter explained how the results presented in the papers contributed to address the research questions identified in Chapter 1. The above findings and contributions will be considered in detail in the next chapter which, as illustrated below, will focus on discussing the research results by considering their implications for research, theory and practice, by identifying their limitations, and by providing recommendations for future investigations.

Figure 17. Progress Tracker – Moving to Chapter 5



# 5. Discussion and conclusion

The aim of this chapter is to discuss the results presented in Chapter 4 and, in particular, to reflect on the implications of these results from a methodological and theoretical perspective. This chapter also draws attention to a number of research constraints and limitations, and suggests areas of future research in order to further develop and deepen the analysis of expertise in the shiphandling context.

## 5.1. Contributions of the research

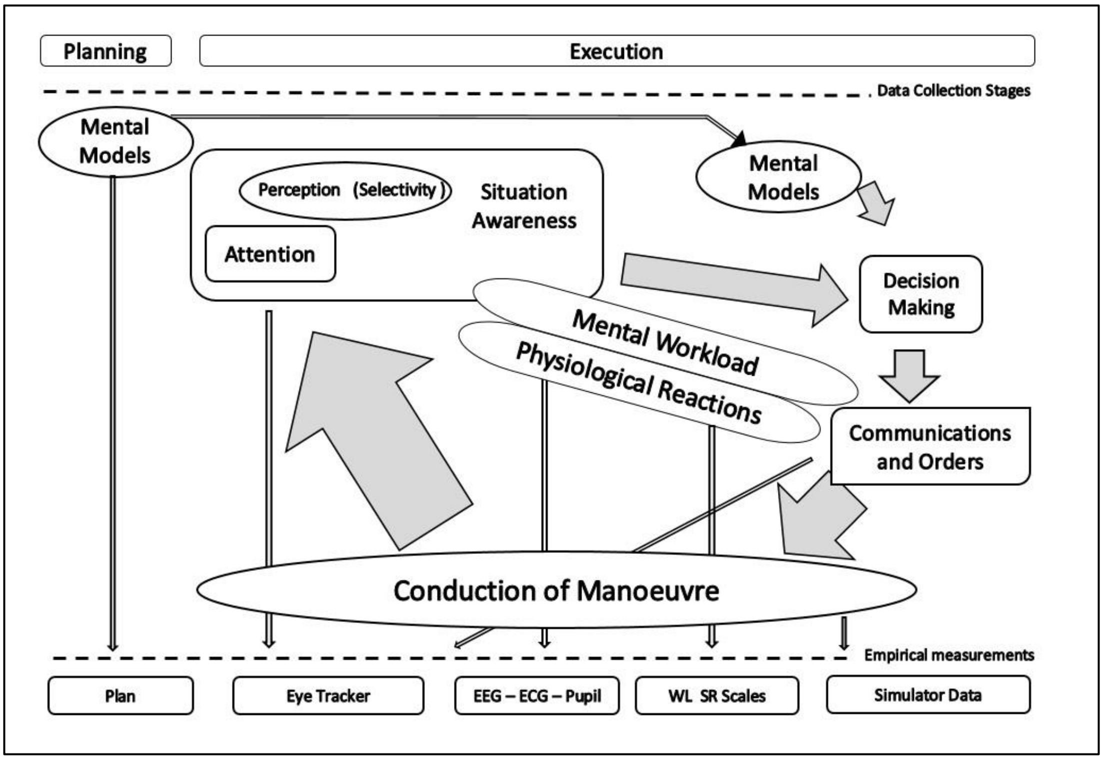
The contributions presented in the previous chapter are threefold and relate to (1) the use of mental models as benchmark to evaluate performance while executing shiphandling manoeuvres, (2) the measurement of mental workload in different shiphandling conditions, and (3) the use of gaze and audio recordings to gain insight in the acquisition and maintenance of situation awareness. These results aim to further our knowledge in the field of applied expertise, as identified in the industrial domain of the maritime industry and the specific knowledge and skills associated with shiphandling.

This research started its contribution by providing an example of a standardised methodology that unpacked the complexity of shiphandling in some of its simpler components (see the discussion on paper I in section 1.1.1). The study showed how, using different devices, several measures could be profitably obtained and translated in meaningful dependent variables. These variables were processed and statistically analysed, providing an empirical basis onto which different theoretical constructs were explored and discussed. With reference to these specific contributions a more detailed explanation will be provided in the following paragraphs. To further support the discussion of those constructs in this chapter, it is possible to refer to Figure 18, which is a simplified version of the theoretical model initially introduced with Figure 2 in paragraph 1.2. It is different from Figure 2 in that Figure 18 reports only those theoretical constructs that were specifically addressed in the empirical research of this thesis.

Briefly, Figure 18 describes how, in this research, shiphandling expertise was captured analysing different elements of a combination of linear (from planning to execution) and circular (the maintenance of components of expertise) processes. The start is represented by the planning stage. This was when the manoeuvre was presented to the pilot to be analysed. Pilots knowledge and experience helped them to understand the implications of manoeuvring conditions (ship type, environmental conditions, etc..). This understanding allowed pilots to select their preferred strategy to perform the manoeuvre. This strategy was practically condensed in the creation of a detailed manoeuvring plan (DMP), and theoretically was encapsulated in pilots' mental model of how the manoeuvre was expected to develop. That mental model was then brought into the actual execution of the manoeuvre. Such mental model provided the basis to compare the outcome of the execution through the iterative cycle that is proposed in Figure 18. Pilots' understanding of the outcome of the execution was considered to be pilots' Situation Awareness (SA). Situation was gained and maintained through and active search in the environment carried out by focused attention, as involved in perception processes (adopting the perceptual cycle interpretation of SA). Specific cues were searched, found and compared against held mental models. The comparison between actual state (in the form of continuously acquired situation awareness) and desired state (as represented in the plan, considered a proxy for the mental model) fed decision making.

Through decision making, pilots assessed consistencies or discrepancies between perceived and desired states, and acted accordingly. Communications and orders were pilots' tools to act on surrounding reality. Based on pilots orders, that affected ship conduction, new outcomes were obtained. Based on these new outcomes the whole process repeated itself as described. How this study addressed all these elements constituting this cyclical process, and the results obtained, will be better discussed in the next paragraphs.

Figure 18. Model of Shiphandling Expertise adopted in the research - Simplified



### 5.1.1. Expertise and its relationship with Planning and Performance

When assessing the execution of a task, an evaluation can be obtained comparing the recorded outcome against an ideal result. Shooting a target, could provide a clear example. The perfect shot is the one that hits the centre of the target. The more a shot departs from the centre, the lower the points that the shooter receives. Using this analogy, an expert hits the centre of the target more frequently than someone with less expertise. In paragraph 2.1.1 it was highlighted how expertise is domain specific. To understand the essence of expert performance in a specific domain, researchers have created standardized representative tasks that could be presented to a group and then identified those skills and results that best discriminated experts (K. Anders Ericsson, 2006a). To our knowledge this standardized representative task does not exist for handling large ships.

The first step in this study was then to create suitable tests or manoeuvres. Referring back to the example of the shooter, the first question then was: would it be possible to create an analogous “target”, suitable to highlight elements of shiphandling expertise? Fortunately, as detailed in paragraph 3.2, to answer to this first question, it was possible to rely on the use of a full mission bridge simulator. Similarly to a laboratory, a full mission bridge simulator offered the possibility to replicate exactly the same experimental conditions. More than a laboratory, though, the simulator provided such a high level of ecological validity, that the experimental setting could be considered similar to that of a field study. The next question then was: in order to clearly identify components of expertise, would it possible to define the perfect execution of a manoeuvre (if “the best manoeuvre” could be defined, would it be the fastest or perhaps the shortest..?)?

In reality, shiphandlers know that there is nothing more evanescent than the definition of a “perfect manoeuvre”. In the real world every manoeuvre is different, even though two manoeuvres might be conducted with the same ship, in the same port and to the same berth. The reasons are many: every ship has her own manoeuvrability that depends on hull shape, propulsion, steering, loading conditions etc... The environment plays an extremely important role. Different wind, tidal and current conditions will deeply influence the way the same vessel will respond. In a real port, all these “parameters” constantly change, making shiphandling every time a different exercise. In addition, personal preferences of the shiphandler and choices to adopt different manoeuvring techniques play their part. So, how would it be possible to tame such variability into a standardised evaluation?

As detailed in the introduction, in this research participants that were selected were considered “experts” in shiphandling, based on their years of experience and current employment as marine pilots. As experts it was reasonable to assume that they understood the implications of the manoeuvring conditions provided to them. It was assumed that, taking into account those implications, they were able to provide a manoeuvring plan, a strategy, capable to bring safely and successfully their ship alongside the assigned berth. As detailed in paragraph 3.4.1 the first stage of the research was to specifically ask the participants to verbalise their intentions, once they were made aware of what the manoeuvring conditions were. At this stage, through a direct exchange between the researcher and the participants, manoeuvring intentions were translated into numerical quantities (see Appendix 6) that could be compared later on with the execution. Using the example of the shooter, this "Detailed Manoeuvring Plan" served as the "target", as the "ideal manoeuvre" against which it was possible to make the evaluation. In the context of this study, the more the execution matched the plan the more the shiphandler was able to prove his/her competence. The novelty of this approach was then to evaluate the participants on two of the most important aspects of expertise: capability to plan and capability to execute. Matching plan with execution may not encompass all the possible manifestations of shiphandling expertise. Nevertheless, those two activities both imply in-depth knowledge, to forecast what to expect in the manoeuvre and to timely and accurately interact with available resources and means. Lacking in any of the two would have proved succeeding in the simulator very difficult.

As already encountered in paragraph 2.1.8, where theories of expert Decision Making were summarised (see Table 4 (G Klein et al., 2006; Salas et al., 2010)), one of the most important characteristics of experts was that they own accurate internal representations of how things work in their domain of practice (Rouse & Morris, 1986). Experts' intuition utilises situation assessment and problem representation, which includes maintaining an understanding of the entire picture (Mica R Endsley, 1995; Flin et al., 1996; Mosier, 1991; Randel et al., 1996). Experts engage in problem detection, identification, anticipatory thinking, forming of explanations, identifying explanations, discovering inadequacies in initial explanations, and projecting the future (Gary Klein et al., 2007; Gary A Klein, 1993; Weick, 1993, 1995). The detailed manoeuvring plan (see paragraph 3.4.1) obtained from the pilots before the manoeuvres were conducted in the simulator, captured and quantified all those elements. Referring back to section 2.2, the DMP (Detailed Manoeuvring Plan) acted as the practical translation of pilots' mental models, meaning the "mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future states" (Rouse & Morris, 1986). This was the first component of expertise that was considered in paper II.

As it can be noticed in Figure 18, the mental models were formed during the planning stage, and were then transferred to the execution stage. In the execution stage those mental models, were compared to reality, developed and maintained as pilots' situation awareness. Mismatches between mental models (expectations) and situation awareness (perceived reality) fed Decision Making. Communications and orders translated into action decisions taken by pilots aiming to rectify perceived those mismatches.

At this point, moving into the conduction of the manoeuvre, it was considered the second fundamental component of expertise considered in paper II: accurate performance when experts are acting in their domain (G Klein et al., 2006). This was the aspect that was specifically explored with the detailed analysis of the execution of the manoeuvres. Smooth and efficient actions were expected, while achieving the desired outcomes (as expressed in the DMP) (Posner & Snyder, 2004).

Those two fundamental aspects of expertise, planning and executing capabilities, were compared against each other to obtain an overall evaluation. If pilots were not capable to foresee the implications of the proposed manoeuvring conditions, they would have encountered serious difficulties to put their intentions into practise. On the other hand, given the accuracy of their plan, to complete their manoeuvres they had also to follow up in the simulator, demonstrating competent vessel conduction. The dependent variables that were developed in this research, were specifically designed to target the discrepancy between the plan and the actual execution. Those variables were comparing a “forecasted quantity” as expressed in the DMP, against an “actual quantity”, as recorded during the execution in the simulator. These elements are shown at the bottom of Figure 18 (analytical comparison of the “plan” against the “simulator data”).

As outlined in paragraph 4.1.2 several results were obtained. Among those results, the XTD (the dependent variable measuring the distance between the expected position against the actual vessel position during the manoeuvred) clearly indicated how higher scores were recorded during the swing. The swing was the phase when the vessel had to be rotated 180 degrees. In this phase, pilots were statistically able to remain within 100 meters of their intended position 80% of the time. This information becomes crucial when swinging vessels in constrained waters, where 100 meters could make the difference between a safe manoeuvre or an incident. The contribution of this study, though, was not simply in the specific quantification of the 100 meters (which was dependant, of course, on the experimental conditions). The contribution was also in how the quantification was carried out, providing an example of how, when intentions differ from reality, this can be accounted for, especially when conducting manoeuvres in a real port. Another example was obtained from the analysis of the difference between the intended use of the main engine and its actual use (independent variable EngEst). Pilots expected to use the propulsion much less than what experienced in the simulation. In the swing the difference between the planned and the effective use, reached values of 50%, when the engine was already working up to 80% of its maximum power. Pilots’ plans did not consider to use the main engine that much, nor so close to its maximum availability. These results showed how it is possible to identify and quantify discrepancies that may depend either on lack of initial understanding or limitations in execution skills (or both). Either ways, similar outcomes would help to identify were the manoeuvres, in particular conditions, could require a different approach or more training (or both) to increase safety margins.

This approach provided the advantage of detaching the assessment from port or environment specificity. The manoeuvres chosen in this study were but a few of the many that could be developed to replicate and test specific conditions of interest. Nevertheless, the way to assess the outcome would be exactly the same: asking in advance what is expected in the manoeuvre according to the specific conditions, and then compare and assess the actual execution against such prediction. The practical achievement of this analytical comparison, as detailed in paper II (see section 6.2), can be applied for any intended or adopted manoeuvre in any specific port. In a more realistic environment, where manoeuvring context would dynamically develop due to, for example, changes in environmental conditions or the occurrence of unexpected events, this approach would still be practicable. Should an unexpected event happen (i.e. a mechanical failure), pilots would still need to assess the situation and consequently take action, keeping aware the bridge team about their understanding of the implications. This is exactly the process that should occur in emergencies and the application of contingency plans. It naturally follows then that the assessment of Pilot's reactions in emergencies and their application of contingency plans would further support the assessment of their expertise.

The capability to predict the expected outcome and translate it into practise is fundamental for expert shiphandling. Every day Pilots commit to take vessels in and out of ports, based on the assumption that they will be able to maintain control of the vessel within specific safety boundaries. Their timely appreciation of significant changes in shiphandling conditions (vessel characteristics, wind, current, tide etc..) and their understanding of the consequent implications is what (or at least partially) can define their expertise and keep ships and ports safe.

The results provided and discussed in paper II (See section 4.1.2), illustrated how theoretical elements of expertise were unpacked, measured and quantified. But how does this impact the reality of the shipping industry? Based on the manoeuvres adopted, this research was able to define differences and probabilities within which the execution was able to match the planning in the group of participants. Similar and more tailored work can be carried out for specific manoeuvres, in specific ports with different groups of pilots. Achieving such empirical quantification becomes extremely important in the conduction of real world activities. Such evaluation is the main and most difficult exercise in the conduction of risk assessments that would decide the suitability (or not) of certain operations. Deciding to carry out certain port operations (ship manoeuvres) without a proper and thorough appreciations of the risks involved, is likely to increase the probability of facing the sorts of low probability, high severity incidents that must be avoided. This research also has direct implications on broader safety management issues, and although not the central issues in this thesis, and not an exhaustive list, these are briefly mentioned below:

**Achieving High Reliability** (highly reliable organisation theory (Roberts, 1990; Weick & Roberts, 1993)) - As Beyea describes in her work, Highly Reliable Organizations (HROs) are those that are known to be complex and risky, but still safe and effective, even at high levels of operational performance/demand. High reliability organizations, committing to safety, value teamwork and nourish a culture of continuous learning and improvement and redundancy in safety measures and personnel. One of the primary drivers of those organizations is that errors are important opportunities to learn, and when they occur, knowledge gained from them helps preventing similar events from occurring in the future. This is achieved by identifying how mistakes were made, examining errors to determine their root cause instead of blaming individuals (Beyea, 2005). Planning is instrumental to high reliability organizations, since it condenses, incorporates and applies all the lessons learnt from previous experiences. Not only, planning provides the inclusive framework within which individuals can share their contribution and coordinate their efforts with other members of the team. The contribution this thesis makes in the context of planning can therefore support higher levels of reliability.

**Engineering Organisational Resilience** - Engineering resilience in organizations aims to enhance the ability to create processes that are robust yet flexible, proactively using resources when disruptions may occur. In Resilience Engineering, failures are not interpreted as a breakdown or malfunctioning of normal functions, but rather shortcomings in the adaptations needed to face real world complexity. It is assumed that individuals and organisations must always adjust their performance to the current conditions. Since time and resources are finite it is inevitable that those adjustments are approximate. The goal then becomes to anticipate the changing risk before damage occurs, with failure simply considered as the temporary or permanent lack of that anticipation. The way to build such capacity is understanding how to create adaptive margins into systems, able to anticipate and absorb pressures, variations and disruptions. Here are some fundamental traits of organizations and individuals applying resilient engineering. Present success does not guarantee future safety. Past results are not enough to rely on their adaptive strategies for future success. Risk is always consciously considered, even when everything looks safe, since the idea of what is risky may have become old or wrong. Doubt is welcome as well as opinions from minorities, to maintain an open-mind and remain sensitive to changes (Sidney Dekker & Cook, 2008). In this context, planning can offer the opportunity to raise concerns and consider alternative options. Plans would include contingencies, accounting for foreseeable disruptions, increasing the probability of finding a way out. By providing a more detailed understanding of the nature of marine pilotage expertise, the knowledge developed in this thesis can contribute to more resilient shiphandling in several ways:

- Obtaining more precise measurement of performance in the simulator,
- Identifying mismatches between mental models and executed tasks, and also;
- Identifying issues in how pilots maintain their perceptual cycle over the course of a manoeuvre.

**Managing Automation Challenges -** As documented in the literature (Billings, 1991; Neville Moray, 1986; C.D. Wickens, 1984) adoption of automation may expose operators and systems to what has been referred as the "out-of-the-loop" performance problem (Mica R Endsley & Kiris, 1995). This problem refers to operators of automated systems and their limitations in the ability to take over manual operations in the event of automation failure. There are many contributing factors to this phenomenon. Some of them are: a possible loss of skills arising from complacency, the shift from active to passive information processing and the change in feedback modalities made available to the operator. Endsley and Kiris (1995) point out how a lower involvement of operator control when interacting with automation is a major contributor to the loss of SA, although this concept of SA "loss" is acknowledged as being based on a somewhat circular argument. So how would planning be involved in the mitigation of such risk? The results of this thesis suggest that proper planning would include when automation would be allowed (and at what level) or in which context manual conduction should be resumed. An example could be provided referring to the use of autopilots on board of vessels. Integrated navigation systems or dynamic positioning systems are technologies that allow operators to set a track on a charting system and have the vessel automatically following that track, without operator intervention. The thorough understanding by operators of how such technology works, with its strengths and its weaknesses, becomes fundamental to understand the limits of applicability and use of those features. Safe manoeuvres in ports, depending on circumstances, may or may not suggest the use of certain technology. Planning would be the stage at which to consider the suitability of automation options. During planning would be also the time when to state how changes in levels of automation should occur, in order to smoothen otherwise traumatic transitions from automatic to manual control. Further, the measurement of workload and gaze has implications for the understanding of this type of automation-related problem – how the pilot adjusts their attention once they are required to take manual control, and how their workload changes during that time.

**Port team resource management** - In a recent report, Goodfellow (2008) argues about the eroding competence of crews. Risks rising from reduced crew competencies, require proper mitigation using all available means that a port can deploy. These means include and are not limited to VTS (Vessel traffic services), tugs, linelaunches, linesmen etc.. The master-pilot exchange briefing (2007; Wild & Constable, 2013) is meant to offer to the captain (bridge team) and the pilots an opportunity to exchange all the information relevant to the upcoming manoeuvring operations. The sharing of that information with the port team, to whatever extent is reasonable, supports more effective and resilient operations, identifying and assigning clear goals and tasks to those involved in the process. This thesis makes a contribution to this issue by identifying measures that can profitably be used to optimise port team performance. For example, measuring gaze (as a proxy for attention) and workload across a team could be used to optimise workload and the perceptual cycle of all participants – rather than the current situation which has tug masters and VTS Operators playing much more passive roles.

### 5.1.2. Expertise and Mental Workload. Measurements and implications

As summarized in paragraph 2.1.3, Feltovich, Prietula et al. (2006), explain how in the traditional theory of expertise (P.M. Fitts & Posner, 1967), skill acquisition is progressive. Initially there is the closely controlled acquisition of a novel cognitive task and then, with subsequent practice (and even more so, with deliberate practice (K.A. Ericsson, 1996; K. Anders Ericsson, 2002; K. Anders Ericsson et al., 1993)), actions become smoother and more efficient, till the stage when performance is achieved with a minimal effort, running essentially automatically, without active cognitive control (Posner & Snyder, 2004). This is when more processes can run in parallel. Thanks to automaticity, experts can show an important reduction in cognitive demand while performing in their own field of expertise (Shiffrin & Schneider, 1977). From these considerations, becomes particularly evident the importance of the relationship between workload and expertise. More specifically, as described in paragraph 2.3, in the *requirements resources interaction* approach, mental workload is defined in terms of interaction between human capabilities or resources and task requirements (P. Hancock & Chignell, 1986; Wieland-Eckelmann, 1992). Certain authors assume that workload is the effect of task demands on a single, undifferentiated pool of resources (Gopher & Braune, 1984). Recent theories though suggest that operator's resources are engaged differently and even independently, depending on the type of task (Hollands & Wickens, 1999; Jex, 1988). Despite of the different positions on the matter, whether task demands would drain operators' resources from a single pool or from several components, what clearly appears is that operators' resources are limited. Therefore, if operators are faced with excessive levels of task demands, they might become incapable to cope, incurring in excessive levels of workload. This phenomenon is what has been defined as "crossing red lines" of performance (Grier et al., 2008). To further investigate this issue of an upper "redline" for task demands, in paper III (see section 6.3) it was mentioned Wickens' model of multiple resources and performance prediction (Christopher D Wickens, 2002). The model defines separate resources: three dimensions associated with different stages of processing, different "codes" of processing (e.g., visual and language are "coded" differently for processing) and different modalities (that indicates auditory perception uses different resources than does visual perception). The value of the model is in "predicting relative differences in multi-tasking between different conditions" ((Christopher D Wickens, 2008) p.452).

While this research did not specifically apply the method identified in Horrey and Wickens (2003), the attributes of the model were useful to comment the results presented and discussed in paragraph 4.1.3. In summary, resource supply and demand for marine pilots is an issue that has attracted debate within the maritime industry for a number of reasons. There has been some argument as to whether the embarkation of a pilot increases a bridge team by one or reduces the team to one. The erosion of the competency of seafarers to support a pilot during the manoeuvre has been questioned (Goodfellow, 2008). Tug masters have traditionally followed verbal orders rather than using their expertise to interact as part of a team. The Vessel Traffic Service (VTS) has almost unilaterally played a minor role in the process, even though it is becoming clearer the importance of a more active contribution (de Vries, 2015).Each of these elements (ships' bridge teams, tug masters, Vessel Traffic Service operators) might provide significant opportunities in the external resource supply side to reduce the task demands imposed on marine pilots, especially during manoeuvres in critical conditions. In light of all those considerations, and with the aim to further support the research in the direction of identifying critical levels or "red lines", one of the contributions in the thesis was to collect and compare different concurrent measurements (direct and indirect) of mental workload in different experimental conditions, but as close as possible to a realistic environment.

With further reference the concept of "red lines" (Grier et al., 2008) and its characterisation as the level of workload beyond which performance would be unacceptable, we can look now at section 3.5.11 (Missing Data) with different eyes. In this section it was reported that during the execution of the forty manoeuvres, a total of four crashes were recorded, as well as three impacts. Of the seven events mentioned, six were recorded during the difficult manoeuvres. The number of events was not statistically significant (chi-square test on the factors port and difficulty) considering the number of manoeuvres. Those results, though, might help to direct the attention to the differences in the levels of difficulty adopted, that were able to induce the increment of events in the more difficult manoeuvres. Table 10 in section 3.3.2, describing the two levels of difficulty, could provide a reference, for future research, to better define those manoeuvring conditions able to reach those "red lines" of workload.

Referring to Figure 18, mental workload was placed in the middle of the entire cyclical process including: conducting the vessel (Conduction of Manoeuvre), comparing expectations (Mental Models) with actual situation (Situation Awareness) and taking decisions (Decision Making) about corrective actions (Communications and Orders). In this representation, mental workload is accounting for the "effort" required to complete the cycle. Paper IV also referred to a cyclical process, the Smith and Hancock's Perceptual Cycle Model (1995), which describes how cyclical interactions between the shiphandlers and the external environment support the maintenance of SA. Results were provided in section 4.1.4, but a further discussion specifically on this topic will follow in the next paragraph. Figure 18 also links mental workload with physiological reactions. At the bottom of the figure, it is possible to see the measurements collected in the study related to mental workload: EEG, ECG and pupil dilation, as well as self-reported scales (NASA TLX and a simple subjective assessment).

The choice of certain physiological measures, intended as proxies for mental workload, was driven by the literature review summarised in section 2.3.3. In this sections the advantages and disadvantages of measuring: EEG, ECG, respiratory frequency and pupil dilation were explored. These types of measures are based on the premise that workload will induce bodily changes. In general, these measures are less convenient to use than performance and subjective measures, but they can provide important advantages, such as being unobtrusive (Eric Farmer, 2003). One of the contributions that this research provided was to demonstrate if and how those measurements could be effectively collected in an experimental setting as close as possible to real working conditions. With the aim of applying such approach in real world scenarios eventually, the selected devices were chosen for their specific characteristics of being light weight, wireless and relatively simple to setup (see paragraph "Equipment and Questionnaires" in section 3.4.2.). The simulator environment was able to demonstrate the advantages and / or the difficulties related to each type of device and its measurements.

The use of a t-shirt with electrodes embedded into the fabric to collect the ECG signal, turned out to be a reliable data collection tool. The ECG signal was both stored locally on an SD memory while broadcasted to a laptop via a Bluetooth connection, providing redundancy and resiliency in the data collection. Overall, the device proved to be reliable. The signals collected were strong (mV) and clear, allowing confidence in the outcome presented in this study. Based on these results, it is anticipated that these measurements and similar recording devices could be profitably adopted in real-world contexts. Significant correlations obtained between concurrent measurements of mental workload and heart rate, suggest that this physiological measure should be positively considered in future research on the topic. Strong correlation between an unobtrusive physiological measure and mental workload is the prerequisite to monitor and assess the latter simply referring to the former. The advantages would be several: a physiological measure can be continuously monitored and recorded without requiring direct involvement of the subject, physiological measures can hardly be counterfeited. Identifying that exceeding certain levels of workload has serious implications for the safety of the operations carried out by pilots, would inevitably underline the importance of monitoring those levels. The choice of a suitable physiological proxy, for the advantages above described, could provide a valid option. In addition, thanks to the increasing development of accessible technology able to easily provide certain measures (i.e. fitbits and similar devices), long term longitudinal studies can be performed, helping to better understand how workload may have an impact on this specific category of workers.

The collection of the EEG signals was found to be far more problematic. EEG signals are weak (μV) and very sensitive to artefacts (muscles contractions, eye movements, changes in electrode skin contact impedance). As highlighted in paper III (key results provided in section 4.1.3.), those were the reasons why the outcomes obtained from the associated dependent variables, recorded in a dynamic experimental context such as an actively operating pilot, could not be considered reliable. Unless more resilient and compact devices become available on the market, the advice is against the use of such measurement as a suitable indirect indicator of mental workload in the context of shiphandling.

Eye trackers were primarily used to collect gaze distribution and audio recording. The use of gaze and audio data will be specifically discussed in the following paragraph dedicated to the description of its relationship with pilots' situation awareness. Eye trackers were able to record also pupil dilation. Pupil dilation did offer some indications that such measurement would be sensitive to mental workload. Similarly to EEG, though, pupil dilation may be very sensitive to artefacts, the most important of all could be simply induced by a change in illumination conditions. In the real world, such event wouldn't be rare at all, considering for example the need that a pilot has to move from inside a ship bridge, to (possibly) an open bridge wing etc.. This vulnerability would suggest extreme caution in the adoption of such measurement in real operating conditions.

To support and validate results obtained in the physiological measurements, several subjective measures were collected. This type of measurements relied on the assumption that the operator was the best evaluator of the mental effort he/she was experiencing while performing the required task. The advantages and limitations of this type of measurements were previously discussed in paragraph 2.3.2.

In this research, the use of a Likert scale to report levels of experienced workload, offered a very effective tool in experimental condition and in the research context. The collection of this data obviously relied on an honest feedback provided by the participants at any point in time. The reliability and the suitability of such measurement in a real manoeuvre would be obviously dependent on the specific context into which such technique would be adopted. Similar considerations would apply also to the use of the NASA TLX (Task Load Index). In addition, similar questionnaires wouldn't be able to continuously capture pilots level of self assessment during the manoeuvre, but only provide an overall evaluation of an entire manoeuvre. In the context of this research though, those measurements were extremely helpful, since they were used as the primary mean through which workload was assessed in its intensity (Likert scale) and in its intensity and nature (NASA TLX). Those were actually the measurements used to validate the physiological measurements. The Likert scale in fact was the primary tool used to evaluate mental workload throughout the manoeuvres. The NASA TLX, on the other hand, in addition to provide an overall evaluation of each manoeuvre, through its subscales, helped us to better define and support the definition of the nature of the experienced workload. Results reported in paper III clearly showed how the highest scores in the NASA TLX were recorded in the subscale 'mental demand'.

Physiological measures do have their own limitations. The adoption of physiological measures implies having to address, in the data processing, the intrinsic variability of individual differences. Measures such as heart rate or pupil dilation can be compared between individuals only after having performed suitable score transformations. In this research it was demonstrated how the adoption of quantile normalization, as detailed in section 3.5.6., offered a suitable and effective approach. The downside of score transformation, though, is that it can be performed only once all the scores have been collected. This may limit the possibility to monitor those measurements in real time, while the manoeuvre is happening.

Mindful of the relationship between expertise and workload, as summarised in section 2.3., this research was able to demonstrate how different experimental conditions could elicit significantly different levels of workload, which, in turn, may correlate with critical reductions in performance. While there were many statistically significant results identified, these are by no means conclusive. Still, the herein introduced methodology does provide a significant starting point from which further studies could be conducted.

### 5.1.3. Situation Awareness and Decision Making in Pilotage

Referring to Figure 18 it is possible to notice how the model adopted in this study was cyclical: shiphandling was considered a constant reiteration of steps which represented fundamental psychological constructs and processes. In this paragraph it will be discussed how the use of eye trackers provided valuable insights into some of these elements. In this research, the use of eye trackers was dedicated to the collection of two types of data: eye gazes and voice recordings. Due to space constrains, paper IV could only include results obtained from the analysis of gaze data. As shown at the bottom of

Figure 18, eye gazes and voice recordings related to two very different elements of the shiphandling cycle. Gaze data was specifically collected to closely monitor how pilots' attention was shifted throughout the manoeuvre. Being able to monitor pilots' attention, enabled to understand what pilots considered (or not) relevant at any particular point in time, to gain and maintain their situation awareness.

Audio recordings, on the other hand, were able to account for pilots' orders and communications. Orders were the means through which pilots were able to act on the environment. The issue of an order implied that a mismatch was perceived between the actual state (as internally owned by pilots as situation awareness) and a desired state (internally owned as mental model). If pilots decided to act (Decision Making), their order was the mean through which they could act on the environment to rectify the situation. As summarised in paragraph 2.4.2, this process was described in the literature by Smith and Hancock (1995). They used their perceptual cycle model to explain the achievement and maintenance of situation awareness. They postulated that internally held schemata directed a person's interaction with the world. The outcome of the person's interaction then modified the original schemata, which in turn directed further interaction with the world in a cyclical manner (K. Smith & Hancock, 1995).

A key contribution of this research has been to document this process with empirical measures. Specifically referring to Table 5 in section 6.5 (Paper IV – unpublished version), this study highlighted how internally managed principles and concepts of shiphandling (Pilotage Schemata - DMP), in different experimental conditions (Phase Description), directed significantly differently pilots' exploration of the environment (gaze behaviour) and pilots' actions (orders). More specifically, within these schemata, it was shown how:

• Pilots were engaged in goal-directed actions (i.e., Pilots had particular shiphandling priorities at different phases of the manoeuvre. They shifted from monitoring speed and direction, to focusing on position and momentum and then speed and lateral position on closing to the berth)

• Pilot held more generalised relevant information for that action, and this changed with the shiphandling priority. For example in the approach the focus on speed was associated with knowledge about maximum/minimum speeds necessary given environmental and other conditions.

• Pilots directed their attention to very specific and more relevant sources of information to decide how to act.

According to Perceptual Cycle model, also active interaction between pilots and the environment were included. These actions, constantly aiming to reduce the discrepancy between the actual state as perceived through the information gathered, and the one aimed or desired based on the own schemata, were recorded as orders. This study, through the use of eye tracking and other technologies (i.e., simulators) was able to comprehensively map changes in the perceptual cycle of marine pilots. Moreover, differently from other studies conducted with the use of eye trackers (refer to paper IV in section 4.1.4 for more details), the approach of the study suggested the adoption of specifically defined sequences instead of considering simple individual behaviours or areas of interest (such as a particular piece of equipment like a radar). Those sequences were anticipated to address specific tasks (such as monitoring position, direction, speed etc..) and the data analysis targeted when the participants' gaze was directed to objects or sources of information relevant to those specific tasks. For example, it was inferred that the pilot was checking the direction of the vessel, when a sequence including a gaze on the ship's bow, a movement of the gaze to an object on the background (within 30 degrees off the bow) and a check on a heading instrument was completed. The previously described sequence does not, of course, cover all the possible behaviours that may lead to a check of vessel's direction. It is one of the sequences of behaviours that would be able to fulfil such task. Being the first time that such approach was adopted in the shipping industry, there is still not enough information about the validity of those sequences as proxies for pilots' aims or goals. The advantage and the contribution of such approach is that if certain sequences can be proven capable to infer specific checks, consequently they can be adopted to infer covert elements (at present) of decision making processes. The literature on decision making already informed us that experts have a heightened ability to make fine discriminations among different stimuli coming from the environment in comparison with those that are less expert (Gary A Klein & Hoffman, 1993). The use of eye trackers may help to better identify what these stimuli are.

Elements of decision making processes can be additionally inferred through the analysis of pilots' orders and communications. That is when the collection and the analysis of audio recordings becomes relevant and valuable. In paragraph 2.1.8 (Expertise and Decision Making) are described different paradigms of decision making theories and their relationship with expertise. Early Naturalistic Decision Making (NDM) research discovered that experts expend a considerable effort on situation assessment and that successful professional judgment in the field is radically different from prescriptive processes (Raanan Lipshitz et al., 2001). The Recognition-Primed Decision (RPD) model described by Klein (1993), for example, states that when it comes to high-stakes, time-pressured decisions, experts do not use "rational choice" or utility analysis, instead, they rely on their experience, recognizing the situation as typical, as a prototype. This prototype brings what to expect from the situation (expectancies), suitable goals, typical courses of action (COAs). Exploiting this prototypical knowledge the expert doesn't need to go through elaborate analyses. The initial recognition can lead directly to action with no comparison of options. In the specific case of this research, pilots' orders (and communications) clearly state the chosen course of action, based on their recognition of what they observed (gaze patterns).

Experimental conditions can be optimised to elicit specific tasks. The analysis of the execution would support the validity and the adoption of certain sequences or patterns as proxies for those specific shiphandling tasks. Should future research further support the relationship between behavioural sequences (as captured by eye tracking devices) and shiphandling tasks or goals, this would have several beneficial implications:

- It will be possible to infer pilots understanding and goals simply monitoring their gaze and speech behaviour;
- It will be possible to compare behaviours adopted (as dictated by underlying decision making processes) and outcomes (as measured, for example, by the performance variables described in section 3.3.3.), evaluating which constellation of analysis and response (or better say, shiphandling technique..) was able to reach the best outcome.

#### 5.1.4. Other contributions

In paragraph 1.1 (Research Background) it was discussed how traditional assessment, focusing on written or oral examination of knowledge may be effective in assessing ability to memorise knowledge-based components of tasks, however it will not suffice to determine demonstrated skills, unless integrated with performance-based assessments (J. Biggs & Tang, 2010). The expectations of seafarers and employers may be addressed if Maritime Education and Training (MET) implements authentic assessment that require seafarers to emulate task performance at workplace standards in real-world contexts (Ghosh, 2017). This research endorsed and promoted such approach since it empirically explored how elements of shiphandling expertise (thanks to the support of specific measurements) could be identified and studied. Once those elements were acquired, they could be explained and potentially transferred to new shiphandlers. The learning by experience, where a trainee pilot would try a manoeuvre, make mistakes and gain a valuable lesson first hand under the supervision of a mentoring pilot, would probably never be completely substituted. Nevertheless, more effective techniques can be taught from the start, once adequate research was able to further demonstrate what those more effective techniques were. Shiphandling expertise, would not only be based on specific knowledge and experience, but would also involve the description of well detailed processes (with their goals and outcomes) that are effectively carried out during a job.

## 5.2. Constraints and limitations of the research

The constraints and limitations of the research were identified in each of the individual papers that formed this thesis. Nevertheless, in relation to the research as a whole, two constraints and limitations should be mentioned.

The number of participants represented a clear limitation of this study. Nevertheless, pilots spent an average of eight hours in the simulator performing these tasks, and more time completing plans for the four manoeuvres, allowing a deep and detailed data collection. This therefore constituted 80 hours of almost continuous data collection, with many variables being collected at very high sampling rates. In this regard, the data collection sacrificed breadth for depth. It is recognised the value of larger data sets, and suggested that increasing the number of participants in future studies would provide more definitive results in specific manoeuvres. It has to be acknowledged also the complexity and the amount of time required by pilots to unpack their manoeuvring mental model in a more quantifiable form (Detailed Manoeuvre Plan). This is something that diverged a little from how shiphandling operations are normally run. More importantly, plans are just plans – something that can, and in certain circumstances, should be departed from, as the context dictates. DMP were the best proxy, for the purposes of this research, for characterising the mental model of the pilot. DMP are, however, still a proxy and future research might use other techniques to investigate the relationship between pre-held mental models and their use in practice.

As previously mentioned in paragraph 3.5.11, during the simulations not all the manoeuvres were completed by the pilots. A total of four crashes were recorded (three during the manoeuvres in the virtual port and one in the homeport), causing the interruption of the simulation and data collection. All the crashes were experienced during the more difficult manoeuvres. Even though this level of difficulty implied that the safety limits currently adopted in pilots' homeport were exceeded, this experimental condition was introduced to explore results in a situation that even though not present at the beginning, could develop during a real manoeuvre. Even if ports adopt specific environmental criteria to set the safety limits beyond which manoeuvres cannot be performed, it can happen that these limits can be reached and exceeded while a manoeuvre is in progress and cannot, at that point, be interrupted.

## 5.3. Recommendations for future research

This study is a first example of a shiphandling expertise assessment methodology carried out with the integration of different tools and measurements. It was originally tested in a Full mission Bridge Simulator. Future studies should certainly consider adoption and verification of the same methodology in a real environment. An initial step could be for example the introduction on manned models (Hreniuc & Batrinca, 2014).One of the challenges will be to monitor and record physiological responses, to indicate that higher or critical levels of mental workload were reached. This measurement can be useful in operational contexts where change occurs, such as when a new ship type is to commence operations at the port or when the port decides to develop new berths or adjust approaches / channel dimensions. Such measurement can also be useful to designers of equipment to assess the effect of their equipment on tasks, and therefore workload. Future research could also be conducted to study how workload might influence levels of workplace health for marine pilots and other seafarers, considering effects such as fatigue. There is quite possibly a need to assess this acutely (for example over a shift) and chronically (longitudinally over longer periods of time).

With this in mind, some research questions for future studies could be:

*RQ1. What levels of workload are experienced in the real working environment? How do these compare with levels measured in a simulator? Can workload safety limits (redlines) be defined, beyond which operations may be exposed to excessive risks?*

*RQ2. What is the impact of specific levels of workload on pilot's health?*

The use of eye trackers could also be adopted in real scenarios. Through gaze analysis it would be able to obtain information about gaze distribution and fixation, (and therefore attention) specifically with reference to other sources of information (electronic equipment, external visual aids..) that are yet to be considered in this research. In a procedural environment, as shipping can be, it becomes extremely important to understand if certain checks or communications are completed within a certain time or at all. Following pilots' gaze in real time would allow to ascertain what they see, or sometimes most importantly, what they might have missed. The gain of situation awareness starts with the perception of the surrounding environment. Eye trackers could provide an extremely valuable insight of what becomes part of this process. To safely conduct a vessel requires a constant cyclic monitoring of several sources such as radars, ECDIS (Electronic Chart Display and Information System), propulsion and steering repeaters, wind indicators and echo sounders. This information must of course be cross-matched with watching out the windows of the bridge. Eye trackers could help to understand the importance and relevance of all these sources, driving for example equipment and bridge design improvements. The study of eye trackers recordings would be the first necessary step towards unpacking the knowledge required to engineer those improvements. Those improvements would finally aim to facilitate ship operators in their monitoring tasks, reducing their level of workload and increasing their resiliency.

As it was mentioned earlier, in a procedural world, communications are fundamental to exchange important information and to initiate actions. Pilots operate on board mainly explaining to ships' crews the upcoming tasks and what is required from their side. Pilots give conning orders to conduct the vessel and very rarely personally operate ships controls. For these reasons, voice recordings would also be able to provide interesting research material for future studies. In this context also the communications between pilots and any other agent involved in the execution of manoeuvres (Tugs, VTS, linesmen, linelaunches..) should be considered. Not only communication content could be taken into consideration (type and number of orders, communications to port traffic services), but also meta communication aspects could be considered through the use, for example, of spectrum analysis.

Overall, the collection and comparison of data in a real working environment may pose significant challenges. For real-world data collection, it will be critical for instruments to be as resilient and unobtrusive as possible, in order not to distract or interfere with berthing operations. Should that data collection be possible, it could provide a better understanding of normal and abnormal, personal and group response levels, which could help to identify critical operations and levels of performance. Once those levels were identified, they could be exploited as prodromal indicators of the developing of critical conditions. Beyond this, such data collection would provide an understanding of the realism of the simulated environment through a comparative analysis of the data.

The method described in this paper, if systematically adopted, could provide a valid and reliable basis to better develop training and test manoeuvring techniques. This methodology can be adapted to specific contexts and, analysing results, could help to clearly identify optimal ranges of distances, speeds or use of available means, thus allowing the development of safer and more efficient manoeuvres. Using systematic feedbacks once manoeuvres are carried out and recorded in real situations, it could be possible to refine reliability and further validate simulated models.

Real mooring operations, in addition, once recorded, can be grouped in different categories, comparing the conditions encountered, according for example to a "level matrix" similar to the one used for the Simulator assessment as introduced in this thesis. Such approach may open the opportunity for new avenues of research and provide applications that may include: (a) the creation of standardised simulated exercises to select, train, evaluate and certify pilots based on national standards; (b) identification of more realistic construction criteria for actual/future port developments; (c) more reliable port operations safety criteria through more accurate risk assessments. Such effort would require further empirical analysis on data differently sourced. Comparative studies with different groups of shiphandlers at different levels of experience and engaged in different manoeuvres, used as models. This work would help to standardise scales able to better define the dimensions of shiphandling expertise. From those considerations the following research questions were extracted:

*RQ3. Would it possible to develop a battery of standardised manoeuvres that could be used to obtain an internationally endorsed shiphandling score?*

The importance of standardised assessment is obvious and it is proven by the existence of an endless number of examples that are used for several purposes. To report just a few among the most famous: the LSAT (law school admission test - www.lsac.org) required by many law schools as a selection criteria, the IELTS (International English Language Testing System - www.ielts.org) adopted to test English proficiency for studying, migrating and working purposes, and probably the more "akin" to this research, the well-known Driving Test in its multiple nuances. On the line of those examples, the methodology introduced with this research, explained how empirical measurements can be gained in a simulated (or in a real) environment, allowing the first fundamental step in the process of standardization of those results. An examination of the most relevant and useful source of information sought by pilots, depending on context and manoeuvring conditions, can assist designers and manufacturers to optimise equipment designs, and trainers to teach more efficient and appropriate shiphandling techniques. Once sensitive and objective markers were identified, as it was demonstrated in this study, they could be used to monitor execution as possible predictors of future outcomes. Being able to define and monitor meaningful and sensitive behavioural markers, could allow better evaluation of training outcomes, actual performance and, in the future, real time activities in normal working environments. Future studies involving gaze analysis (and therefore attention) specifically with reference to the source of information (electronic equipment, external visual aids..) preferred by pilots (Itoh, Hayashi, Tsukui, & Saito, 1990) could offer an important insight regarding information resource management and shedding preferences once task demand begins to overcome pilot capabilities (M. S. Young et al., 2015). This leaves a final research question:

*RQ4. Would it be possible to automatically analyse data coming from eye trackers, in order to monitor what would be considered an optimal performance and to promptly identify deviations from that performance?*

Considering the decrease in crew numbers, the conduct and the management of a complex engineering system such as a commercial vessel wouldn't be possible without an ever increasing level of automation. The conduct of navigation relies on integrated navigation systems, the propulsion is increasingly supported by unmanned engine rooms, cargo operations are carried out thanks to integrated loading / unloading systems (on board and ashore). One of the immediate implications of the increase of automation is the reduction of ship personnel. Today it is not rare to see on board of a Cape size vessel (300+ meters of length over all) less than 20 crew members. Tasks and roles on board of these personnel have changed, adapting to and accounting for these new working conditions. Technology may relieve personnel from more menial and manual tasks, but still requires significant attention, since the same personnel is involved into monitoring and maintaining those complex systems. As already discussed in section 5.1.1, the adoption of high levels of automation may induce the "out-of-the-loop" performance problem (Mica R Endsley & Kiris, 1995), where operators experience limitations in the ability to take over manual operations in the event of automation failure (Billings, 1991; Mica R Endsley & Kiris, 1995; Neville Moray, 1986; C.D. Wickens, 1984). Another important aspect is also the possibility that automated systems may actually monitor the operator. This is the case for example of eye tracking systems monitoring drivers levels of fatigue (Ji, Zhu, & Lan, 2004; P. Smith, Shah, & da Vitoria Lobo, 2000). The last research questions hints at the possibility that eye trackers could be a useful tool through which it would be possible to study who is in charge of monitoring.

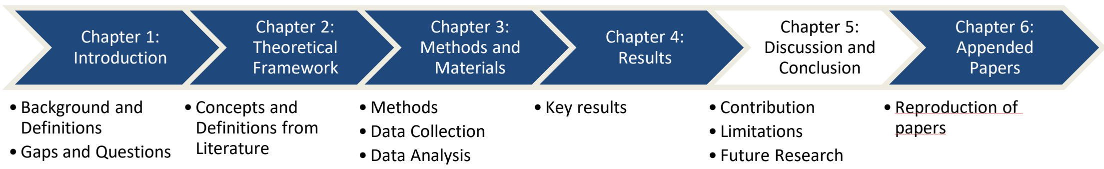
## 5.4. Final observations

This study adds to a growing body of research investigating expertise and its manifestations in different contexts. The research explored a practical approach to translate in quantifiable measures those elements and constructs that have been identified as constituents of expertise in the literature. Even though the study was conducted in a simulated environment, it was conceived to be easily transferrable to a real context. In a real scenario, Portable Pilotage Units (PPUs) and ships' Voyage Data Recorders (VDRs), engine logs, video and audio recordings, can be exploited in order to collect such data on ship bridges.

The use of technology, though, could go much further. Ship manoeuvring requires an understanding and manipulation of complex interactions of masses and forces. It is rare that the effects of these interactions observe linear laws. This makes their appreciation and prediction a considerable task, especially when carried out without the support of appropriate tools and training. This very fact has led other researchers to explore the possibility, through fast time simulations, of making more accurate real time predictions of a vessel's behaviour while manoeuvring (Benedict, 2012). The seamless integration of operators and state of the art technology continues to evolve. Pilots perform their job with different types of vessels, each of them with its unique configuration of bridge equipment and personnel. They need to quickly adapt to the situation, making critical judgements as to the feasibility and the safety of the manoeuvre that they must quickly execute. Technology already provides extremely valid solutions and aids, but it becomes fundamental that those tools are well understood and fully integrated in the job that is carried out every day. Hence the need to have suitable means to guide the training and assess effectiveness and inefficiencies.

It would very difficult to forecast where the maritime industry and the art of shiphandling will be in the next few years. New exciting challenges are waiting to be conquered by the pilots and the seafarers of the future. Unmanned ships could be the next frontier and pilots might not be required to climb the ladder anymore (Yemao, Lundh, & Porathe, 2014). Regardless of the unpredictable changes that the future will bring, if shiphandlers will be still involved, they will still carry with them all those human elements that this research humbly attempted to unpack.

Figure 19. Progress Tracker – Moving to Chapter 6



# 6. APPENDED PAPERS

Chapter 6 reproduces the manuscripts included in the framework of the doctoral project. The five papers have not been rewritten for this thesis. There are, therefore, unavoidable repetitions, especially among the papers as well as between the papers and the thesis chapters. However, in order to ensure the consistency of format, the five papers have been reformatted and the references have been included in the unique list of references at the end of this thesis. With a view to better differentiating the actual papers from the remainder of the thesis, a different format (font, font-size, etc.) has been used for the overall presentation of the papers, i.e. for the headlines, the content, the illustrations, etc.

Table 19. Publication recap

| Paper | Nature | Title | Publication channel | Full-length, double blind review | Status |
|---|---|---|---|---|---|
| **I** | Conceptual | The development of a shiphandling assessment tool (SAT): A methodology and an integrated approach to assess manoeuvring expertise in a full mission bridge simulator. | *Paper presented at the 15th Annual general assembly International Association of Maritime Universities* | Yes | Published |
| **II** | Research | A Comparison of Marine Pilots' Planning and Manoeuvring Skills: Uncovering Mental Models to Assess Shiphandling and Explore Expertise. | *Journal of Navigation* | Yes | Published |
| **III** | Research | Measuring mental workload and physiological reactions in marine pilots: building bridges towards redlines of performance. | *Applied Ergonomics* | Yes | Published |
| **IV** | Research | Interpreting changes in marine pilots’ perceptual cycle through gaze detection patterns. | *Ergonomics* | Yes | Under review |
| **IV AV(1)** | Research | Interpreting changes in marine pilots’ perceptual cycle through gaze detection and speech patterns. | *(1) Additional Version Unpublished* | NO | Unpublished |

Chapter 6.1 has been removed for copyright or proprietary reasons.

It has been published as: Orlandi, L., Brooks, B., Bowles, M. (2014). The development of a shiphandling assessment tool (SAT): A methodology and an integrated approach to assess manoeuvring expertise in a full mission bridge simulator, IAMU AGA 15 Looking Ahead Innovation in Maritime Education, Training and Research, 27 - 30 October, Australian Maritime College, Launceston, Tasmania, pp. 131-140. ISBN 978-0-9806391-4-8 (2014)

## 6.2. Paper II

# A Comparison of Marine Pilots' Planning and Manoeuvring Skills: Uncovering Mental Models to Assess Shiphandling and Explore Expertise

Luca Orlandi, Benjamin Brooks and Marcus Bowles

(University of Tasmania, Australian Maritime College, National Centre for Ports and Shipping, Launceston, Tasmania 7250, Australia)

This paper introduces an assessment methodology that can underpin the objective measurement of shiphandling skills and permit comparative analysis of manoeuvring plans against their execution in a full mission bridge simulator. It was hypothesised that expert shiphandlers would have shown a strong consistency between the initial plan provided and the following execution. Ten marine pilots participated in the study. Their performance was evaluated across several variables using data gathered during the planning and objective measurements completed during the execution on a simulator. A significant capability to match execution against the plan was evidenced by the group of pilots. The mathematical analysis proposed represents an objective approach that can assure a valid and reliable assessment when applied across different contexts and needs such as: selection, training and certification of pilots, port development, optimisation of bridge procedures and improvement of equipment design.

KEYWORDS

1. Shiphandling. 2. Mental models. 3. Ship Simulator. 4. Marine Pilotage.

Submitted: 28 November 2014. Accepted: 10 March 2015. First published online: 17 April 2015.

1. INTRODUCTION. The aim of this study was to better depict the complexity underlying shiphandling expertise in a port environment, with an emphasis on the human element relating to the safety, accuracy and efficacy of ship movements. The study investigated individual competence in a group of marine pilots, to plan and fore-cast future operational needs in different contexts and manoeuvring conditions. Such competence is considered to be of critical importance, since pilots have to decide if a vessel can safely operate in a port, basing their decision on the vessel's manoeuvring characteristics and contingent environmental conditions. Inaccurate evaluations could expose the vessel to critical consequences. The "mental model" concept helps to better contextualise pilots' planning competence in a theoretical background (Mohammed, Ferzandi, & Hamilton, 2010). Mental models have been defined as "mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future states" (Rouse & Morris, 1986). They are generally used to describe a person's mental representations and beliefs of some physical system, with a particular focus on how the individual's interactions with the system lead to the outcome of interest (Hinsz, 1995).

Mental models can also be used to describe abstract dynamics or concepts as deductive reasoning and inference (Aronson, 1997), they could refer to individual or distributed cognitive processes among team members (Banks & Millward, 2000). Effective planning increases shared mental models, allowing team members to better perform during high workload conditions (Stout et al., 1999). Mental models can be seen as knowledge structures which are formed of stored long term static information (Johnson-Laird, 1983) that can be exploited to explain, interact and direct problem solving (Al-Diban & Ifenthaler, 2011). When complex, novel, high risk problems are presented, people rely on mental models as a guide (Mumford et al., 2012) or as a map (Fiol & Huff, 1992). Evaluating how well mental representations are able to forecast future outcomes implies evaluating the prediction validity of the proposed methodology. This approach could improve specific aspects of performance, correcting and refining inaccurate assumptions derived from a partial or erroneous initial understanding. Trainers could adopt different forms of evidence from those they would usually seek to assess performance, modifying learning and assessment events. The current study explored the relationship between pilots' competency to plan several manoeuvres and the execution of those manoeuvres in a simulated environment. This can be seen as the translation in practical terms of their manoeuvring mental models into a simulated "reality". Mental models and outcomes in the simulators were quantified, in order to obtain, through such comparison, a performance measurement. We expected that participant pilots, being "proficient" (Benner, 1982) or "expert" (S. E. Dreyfus & Dreyfus, 1980), were able to formulate plans sufficiently close to execution. In order to contain possible influence of other interfering factors ensuring validity of measurements, participants were also compared with the original company group on several aspects better described in Section 2.1.

## 2. METHODOLOGY

2.1. Participants. The participants of this study were a group of ten marine pilots coming from the same pilot company. They were all males in good health, as required by professional medical standards (AMSA, 2010). At the time of data collection (December 2013) the company had a total of 39 pilots with an average age of 51·2 years at a standard deviation of 7·0 years. All the pilots had an average of 10·8 years of service with the company with a standard deviation of 6·8 years. The group of ten participating pilots were 51·8 years of age on average with a deviation standard of 5·9 years. On average these pilots had been with the company for 10·6 years with a standard deviation of 7·8 years. An Analysis of Variance (ANOVA) for age and service confirmed no significant difference between the participants and the rest of the pilots working for the same company. All the pilots involved in the research had more than ten years of previous experience in pilotage, even if not in the same company.

The experiment was divided into two phases. During Phase 1 participants were required to complete a thorough and comprehensive planning of the manoeuvres that would later be undertaken on the simulator. Phase 2 consisted of observed performance and data collection by the assessor while the pilot executed the previously planned manoeuvres in a simulator. The authors assert that all procedures contributing to this work comply with the ethical standards of our university and the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Table 1. Levels of Difficulty – Adopted in both Ports.

| | Pier - Spatial constraints | Environmental conditions and forces | Vessel characteristics | Tugs | Interactions with traffic | VTS Comms (1) |
|---|---|---|---|---|---|---|
| **Level 0**<br><br>**Easy** | Big Swing Basin<br><br>(3 times Vessel LOA (4)) | Current: 0.7 Knt<br><br>Wind: 15 Knt<br><br>Good Visibility | Single Controllable Pitch Propeller (2)<br><br>Bow Thruster (3) | None | 1 Interacting but not Interfering vessel | General Info<br><br>No reporting Points |
| **Level 1**<br><br>**Difficult** | Small Swing Basin (5)<br><br>(1,5 times Vessel LOA) | Current: 2 Knt<br><br>Wind: 25 Knt<br><br>Poor to no Visibility - Heavy Rain | Single Fixed Pitch Propeller<br><br>No Thrusters | As required by Pilot | 1 Interacting 1 Interfering vessels | General Info and Traffic Advice<br><br>Reporting Points |
| **Notes** | (1) Vessel Traffic Management station present in a port and managing ships via radio communications;<br>(2) Propeller capable to change the water thrust direction changing the angle of the blades instead of direction of rotation;<br>(3) A thruster is a propeller positioned perpendicular to the ship keel axis. Placed on the bow or on the stern, induces transversal / angular motion;<br>(4) Length Over All, maximum length of a vessel;<br>(5) Wider area, within constrained waters, where ships have sufficient room to rotate and revert their direction. | | | | | |

2.2. Phase 1 – Planning. The first phase included the planning of four proposed manoeuvres. Each manoeuvre included the whole process necessary to transfer the ship from a defined initial position to a berth within constrained port waters, with the use of own and/or external means of propulsion (i.e. tugs, when allowed). These four manoeuvres were controlled on three main factors: (a) "port familiarity" (from now on referred as "port"), (b) "difficulty", and (c) "phase". The first factor, "port", took into account whether the manoeuvre was conducted in the participant pilot's homeport (the port where they were regularly working) or in a different port. The other port chosen for the experiment was Vorbasse, a virtual port only present in the simulator software. This port was chosen to avoid any possibility of learning effect associated with previous manoeuvring experience the subjects may have possessed and to provide support for methodology reliability. Vorbasse was also chosen to investigate how "pilots' expertise" could be "bounded", i.e. related to pilots' local knowledge of the port where they normally operate.

The pilots' homeport in the tables and graphs presented will be coded "B", while Vorbasse will be coded "V". For the factor "difficulty" the easy level will be coded "1" while the difficulty level will be coded "2". To control the level of difficulty, specific manoeuvres' parameters were altered as summarised in Table 1.

Level 1 reproduced a comparable level of difficulty of routine operations. Level 2 aimed to engage pilots with a level of difficulty slightly exceeding the safety limits established in the pilots' homeport, without losing construct validity. Each manoeuvre required the pilot to complete a mooring using the side of the ship opposite to the berth position on commencement of the exercise. This implied that for each manoeuvre the ship had to swing (rotate 180°) before she could be moored. Each manoeuvre there-fore developed through three main sections that provided an additional factor for the analysis; (1) the "approach" (from the initial position until the start of the swing), (2) the "swing" (from the start of the swing until the rotation was completed and

Table 2. Proportions between vessels and port dimensions.

| Ship | LOA (m) | Ratio between Ships | Breadth (m) | Disp (ton) |
|---|---|---|---|---|
| Torm Laura (diff Lvl 1) | 183 | 0·7 | 32 | 54925 |
| Arcturus (diff Lvl 2) | 269 | 1·45 | 48 | 143200 |
| Ratio | Torm LOA | Torm Breadth | Arcturus LOA | Arcturus Breadth |
| Basin diameter (470 m) | 2·6 | 14·7 | 1·7 | 9·8 |
| Channel width (300 m) | 1·6 | 9·3 | 1·1 | 6·2 |

stabilised), and (3) the “closing” (from the end of the swing until a defined distance from the berth). In the graphs and tables presented, the phases will be coded: “1” for the approach, “2” for the swing and “3” for the closing.

Manoeuvres were also coupled across the “port” factor (grouped for the same level of difficulty); i.e. the easy manoeuvres in the two ports (as well as the difficult man-oeuvres) were, as much as possible, kept technically similar (e.g., vessels used, distances to be covered, etc.) to promote data baseline formation on pilot performance and assure reliability of the assessment process. Spatial constraints due to port dimensions were purposely maintained to be similar, modifying Vorbasse in order to match homeport dimensions as summarised in Table 2.

Phase 1 required participants to explain extensively how they would have performed the manoeuvres in the simulator, meaning that the plan provided would have been their intended, preferred and expected course of action. Any difference recorded in the following execution would have been considered unexpected and deemed necessary as the best possible option available at the time to maintain the safety of the vessel while achieving the goal of berthing. In order to create and to obtain the record of such ex-planation in a numerical form, a Detailed Manoeuvre Plan (DMP) table was compiled by each participant for each manoeuvre, before performing such a manoeuvre in the simulator. Such a table can be seen as a more detailed version of the routine passage plan normally discussed by pilots and ship masters before a ship enters into a port (Wild & Constable, 2013). The initial material provided by the researcher to pilots included a facsimile of port navigational charts at the appropriate scale for each manoeuvre. Pilots were able to use the charts to sketch the exact expected ship movement and highlight elements of interest. For each sequential position sketched on the charts, the pilot had to forecast in the DMP details such as:

. ship’s speed in knots;

. ship’s main engine power in percentage of maximum power available;

. ship’s bow thruster power (when available) in percentage of maximum power available;

. tug’s force (when available) in percentage of maximum bollard pull available.

Prepared prior to the simulations these plans formed a comparative basis that were used to assess outcomes generated in the simulator. In reality a full mission bridge simulator can record all the previously mentioned parameters (and more) with a high degree of accuracy at several samples per second.

2.3. Phase 2 – Execution. For this research, the Maritime Safety Queensland Simulator located in Brisbane was used (Smartship® Simulator www.smartshipaustralia.com.au). This "Full Mission Bridge" simulator is classified as Class A (NAV) ac-cording to the standards issued by DNV (2011). It is capable of simulating a total shipboard bridge operation situation, including the capability for advanced manoeuvring in restricted waterways. Before the experimental manoeuvres, pilots were required to perform a very simple mooring with a vessel different from those used in the experimental runs. This first manoeuvre was used as a familiarisation run to ensure participants had a standardised level of familiarity with the bridge environment and the navigation equipment available. The manoeuvres planned in Phase 1 were then used in random order to record all the data. To provide realism to the manoeuvres, during their execution, the researcher was present on the simulator bridge and he was generally acting as the ship's Master or the bridge member most suitable for the specific interaction.

Performance outcomes were obtained calculating the following dependent variables:

. XTD – Cross Track Distance: Distance from the intended track as per positions obtained from the planning charts and the ship track provided by the simulator;

. SpdEst – Speed Estimation: Difference between the intended speed over the ground (SOG) as per DMP (expressed in knots) and the recorded speed provided by the simulator.

. EngEst – Engine Power Estimation: Difference between the absolute value of the intended use of engine power as per DMP (expressed in percentage) and the absolute value of the recorded engine power provided by the simulator.

. ThrEst – Bow Thruster Power Estimation: Difference between the absolute value of the intended bow thruster power (expressed in percentage) as per DMP and the recorded absolute value of the bow thruster power provided by the simulator (when applicable).

. Tug(n)Est – Tug Force Estimation: Difference between the absolute value of the forecasted tug's bollard pull as per DMP (expressed in percentage, based on the maximum bollard pull that tugs could provide) and the recorded absolute value of the tug's bollard pull provided by the simulator (when applicable, with (n) differentiating each tug used).

Figure 1 shows two screenshots taken from the simulator interface showing two different manoeuvres (B2 on the left and V1 on the right). It is possible to notice in light grey the outline of two vessels used (Arcturus in the homeport on the left and Torm Laura in Vorbasse on the right). The empty outlines creating the shaded area represent the swept path covered by the vessel during its movement. In the middle of the basins it is possible to note as a segmented line the pilot's intended path from which the XTD was measured.

3. RESULTS. Results against the above parameters were calculated for each manoeuvre completed by a participant. The results obtained were averaged across all participants and within each phase previously identified as "approach", "swing" and "closing". During the simulations not all the runs were completed by the pilots.

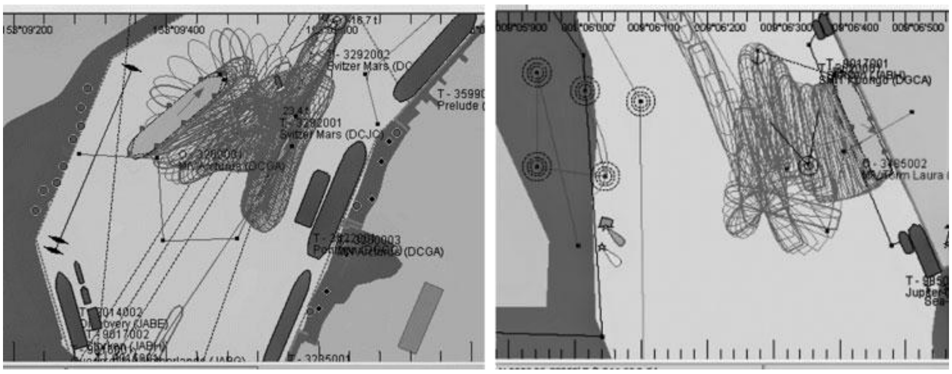


Figure 1. Examples of Manoeuvres as shown by the simulator interface.

Table 3. Summary Table for ANOVA – Significance of Results.

| | Principal Effect | | | Interactions | | | |
|---|---|---|---|---|---|---|---|
| Variable | Difficulty | Port | Phase | Diff * Port | Diff * Phase | Port * Phase | Diff * Port * Phase |
| XTD | | | 0·000 | | | | |
| SpdEst | | 0·044 | | | 0·000 | | |
| EngEst | | | 0·000 | | 0·061 | | |
| ThrEst | N/A(1) | | | | | | |
| Tug1Est | N/A(2) | | 0·001 | | | | |
| Tug2Est | N/A(2) | | | | | | |
| Tug3Est | N/A(2) | | | | | | |

1. Bow Thruster was available only in the easy manoeuvres
2. Tugs were available only in the difficult manoeuvres

A total of four crashes were recorded (three during the manoeuvres in Vorbasse and one in the homeport). A “crash” was an impact or a grounding that required the interruption of the simulation and data collection. All the crashes were experienced during the manoeuvres at Level 2 of difficulty. Three impacts were also experienced (one in Vorbasse with difficulty level 1 and two during the swing in both ports at difficulty Level 2). An “impact” was classified as a contact of the vessel with another ship or port infrastructures that did not impede the continuation of the manoeuvre. Note that Level 2 of difficulty implied that the safety limits currently adopted in pilots’ homeport were exceeded. All the pilots clearly stated during the planning phase that they would have chosen not to conduct those manoeuvres should the stated conditions have occurred in the workplace. For all the variables described in Section 2.3. A Univariate Analysis of Variance (ANOVA) was performed, using the statistical package IBM SPSS (2010), on the factors “difficulty”, “port” and “phase” (as defined in Section 2.2.), obtaining the results reported in Table 3 (showing only significant results with alpha < = 0.05 are reported, all results are reported in the Appendix):

In XTD the comparison of the means was significant only on the factor phase (Sig = 0·000). To specifically identify which of the three phases was significantly

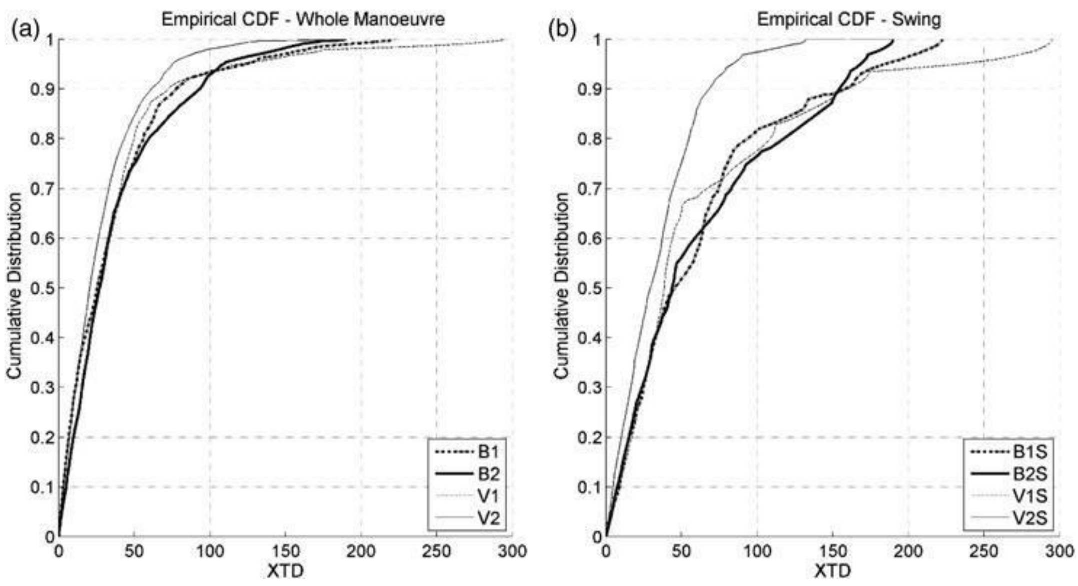



Figure 2. XTD Cumulative distribution function – (a) All manoeuvres – (b) Swing phase.

different from the others, a post hoc analysis using a Tukey's Honestly Significant Difference Test (Tukey HSD) (Abdi & Williams, 2010) was carried out. A significant difference (Sig = 0·04) was found between the phases swing (mean = 55·51) with approach (mean = 26·04) and swing (mean = 55·51) with closing (mean = 38·20). Considering all the manoeuvres performed, pilots showed an averaged XTD of between 21 and 50 metres during the approach and the closing phases and between 38 and 69 metres during the swing. Even though results may suggest that this group of pilots were able to remain, on average, within 40 metres of the intended track across all the exercises, further analysis highlighted other elements of interest. A different perspective was achieved considering the Cumulative Distribution Function (CDF) of the variable XTD. Figure 2(a) shows the cumulative distribution functions of XTD across the whole manoeuvres, while Figure 2(b) graphs those distributions only in the phase swing. Results will be further discussed in Section 4.1.

In the rest of the independent variables that will be explored later, positive values suggest an "overestimation", i.e. the plan provided by a pilot had estimates that exceeded values achieved in the simulator. Conversely, negative values indicate an "underestimation", i.e. the values recorded in the simulator were above those provided by the pilot.

The SpdEst, reported in Figure 3, provided a main significant effect on the port factor (Sig. = 0·044). Pilots showed a deeper underestimation of the vessel's speed in Vorbasse (mean = −0·26) than in their homeport (mean = −0·07). There was also a significant interaction (Sig. = 0·000) between the factor's difficulty per phase. There was an overestimation during the approach of the difficult manoeuvres (B2 and V2; mean = 0·19) compared to the easier ones (B1 and V1; mean = −0·71) in the same phase.

The EngEst, shown in Figure 4, was obtained according to the same rules as the speed calculation. The difference was in the use of the absolute value of the

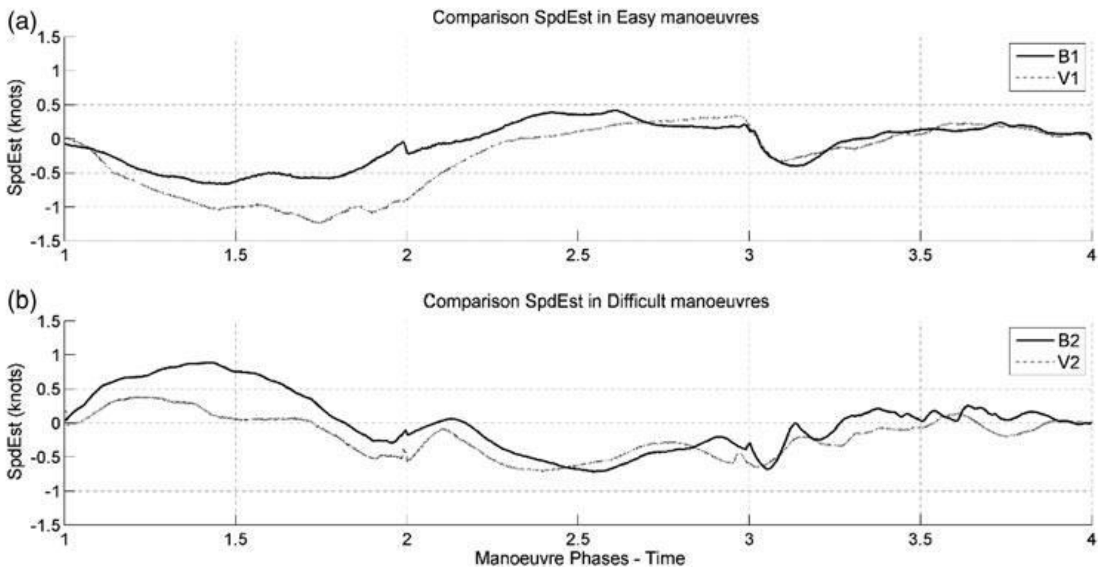



Figure 3. Comparison between SpdEst values in the easy and in the difficult manoeuvres.

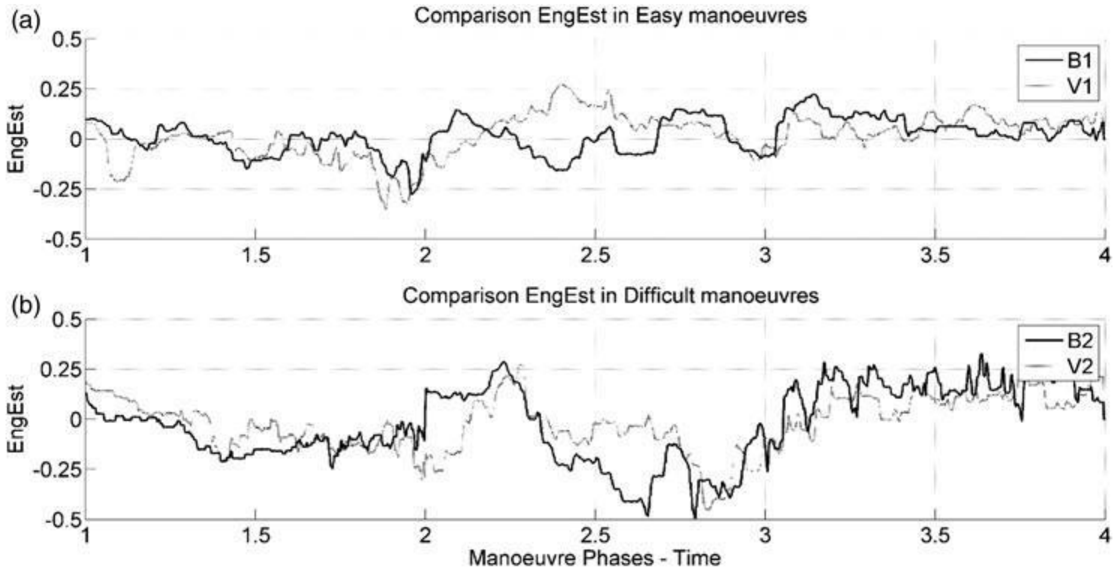



Figure 4. Comparison between EngEst values in the easy and in the difficult manoeuvres.

measurements in the comparison. The engine power used was provided by the simulator with positive or negative values depending on whether the engine was running ahead or astern. Since our focus was on the strain put on the engine in terms of power utilisation and not on the direction induced by the propeller on the water flow, we adopted the absolute value. EngEst, had the main effect (Sig = 0·000) on the factor phase, requiring a post hoc analysis using a Tukey HSD test. Significant differences were found between closing (mean = 0·10) and approach (mean = −0·06) (Sig = 0·001) and then between closing (mean = 0·10) and swing (mean = −0·05) (Sig = 0·006). It also highlighted a marginal interaction (Sig = 0·061) between factors

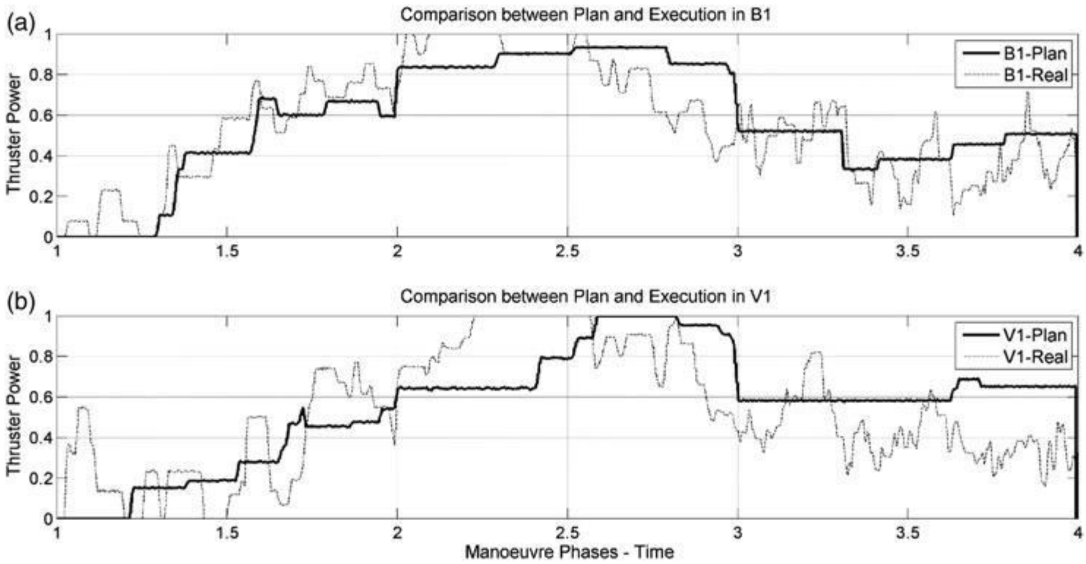



Figure 5. Comparison Plan and Real Use of Bow Thruster in the easy manoeuvres.

difficulty and phase. In the swing an underestimation was recorded during the most difficult manoeuvres (B2 and V2; mean = −0·13) while in the easier ones the use of the engine was overestimated (B1 and V1; mean = 0·35).

Adopting an analogous approach to the one used for the engine estimation, ThrEst was calculated as the difference between the plan and the real use (see Figure 5) of the bow thruster. The bow thruster was available to pilots only in the easier manoeuvres

(Homeport and Vorbasse, level of difficulty 1). A Univariate Analysis of ThrEst showed no significant effects on the factors port and phase.

Tugs were available to pilots only in the most difficult manoeuvres (Homeport and Vorbasse, level of difficulty 2). In order to uniquely identify the tugs throughout the duration of the whole manoeuvre, a number was initially assigned depending on their position around the hull at the very beginning. Tug 1 was the tug made fast on the shoulder of the vessel, Tug 2 on the quarter, while Tug 3 was made fast through the centre lead aft. Even though the disposition of the tugs could have changed throughout the manoeuvre according to pilots' orders, the initial number assigned would have remained the same. As shown in Figure 6(a), only Tug1Est (estimation on Tug 1) reported a significant difference (Sig = 0·001) on the phase factor. A Tukey HSD test showed a significant difference between closing (mean = 0·22) and approach (mean = −0·07) (Sig = 0·008) and significant difference between closing (mean = 0·22) and swing (mean = −0·15) (Sig = 0·001).

Pearson coefficients were calculated for each manoeuvre, to obtain the correlation between the provided values in the DMP and the values recorded by the simulator. The curves representing plan and execution that were compared were obtained through a moving average across pilots. In addition, Pearson correlation coefficients were also used to compare the independent variables' outcomes (with the exception of XTD) across manoeuvres with the same level of difficulty (B1 with V1 and B2 with V2). All the correlations reported in Table 4 provided significant values (alpha <= 0·05) with one exception (see note (3) in Table 4):

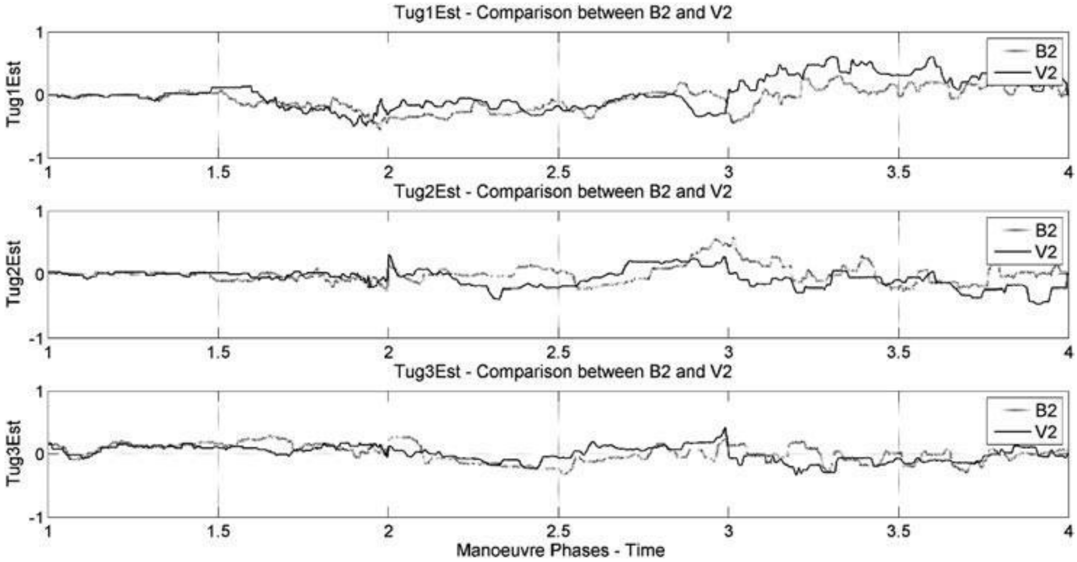



Figure 6. Comparison between Tug1Est Tug2Est and Tug3Est in the difficult manoeuvres.

Table 4. Summary Table for Pearson correlation coefficients.

| DMP vs Simulator Correlations within each manoeuvre | | | | | Correlations between manoeuvres | |
|---|---|---|---|---|---|---|
| B1 | B2 | V1 | V2 | Variable | B1 – V1 | B2 – V2 |
| 0·983 | 0·959 | 0·969 | 0·981 | SpdEst | 0·898 | 0·902 |
| 0·604 | 0·323 | 0·546 | 0·393 | EngEst | 0·267 | 0·652 |
| 0·851 | N/A(1) | 0·672 | N/A(1) | ThrEst | 0·574 | N/A(1) |
| N/A(2) | 0·331 | N/A(2) | 0·040(3) | Tug1Est | N/A(2) | 0·670 |
| N/A(2) | 0·668 | N/A(2) | 0·812 | Tug2Est | N/A(2) | 0·200 |
| N/A(2) | 0·428 | N/A(2) | 0·816 | Tug3Est | N/A(2) | 0·480 |

Bow Thruster was available only in the easy manoeuvres

- Tugs were available only in the difficult manoeuvres
- Not Significant

## 4. DISCUSSION

4.1. XTD – Cross Track Distance. After performing an ANOVA we were able to isolate only one statistically significant result that occurred in the factor phase. The swing was the phase that showed a statistical difference from the approach and closing phases. This empirical result suggests that pilots were generally able to show consistency in their ability to maintain their intended track despite working in different ports and at different levels of difficulty, with a decreased performance only evident when engaged in the swing. This result becomes more evident looking at the graphs reported in Figure 2. Figure 2(a) reports the cumulative distributions of the XTD scores for the whole manoeuvres while Figure 2(b) is specifically for the swing. A CDF obtained from an ensemble of measurements, provides, for any given score, the number of remaining scores that would be lower in value. In a CDF such a number is provided on the Y axis as a fraction of 1, meaning what percentage of scores would be lower than the score chosen on the X axis. That ordinate value, expressed as a fraction, can also be considered a percentage or a probability. In this case, the scores we are referring to are the cross track distances from the intended track (XTD). These curves show that if the pilotage organisation chose 80% as the target probability to remain within a certain distance from the intended track (ordinate 0·8), this requires a distance of 100 metres during the swing, while for the rest of the manoeuvre 50 metres would be sufficient. This implies that if a distance of 75 metres from the intended track was targeted as safe, in the rest of the manoeuvre there would be a less than 20% probability of reaching and exceeding such a distance, while during the swing such probability would increase to around 40%.

In Figure 2(b), it is shown how scores in the easy manoeuvres (B1S and V1S) reported with a dotted line, exceeded in their maximum values the abscissa of the 200 metres, while the most difficult manoeuvres remained below 200 metres. The ex-planation for this counter-intuitive result could reside in the fact that in the easier man-oeuvres the ratio between the dimensions of the vessel used and the dimension of the available swinging basin was more favourable (2·6) compared to the one available for the more difficult manoeuvres (1·7). Pilots were able to exploit more space in the easier manoeuvres (for example to allow more time to reduce speed) while in the more difficult ones a similar range would have resulted in an impact or grounding. It has to be remembered that the scores collected during the swing and during the closing only occurred with those manoeuvres that were successfully completed without crashes.

4.2. SpdEst – Speed Estimation. Results show that pilots estimated the speed in the two ports differently. Pilots underestimated the speed in the port of Vorbasse (−0.26) slightly more than in their homeport (−0·068). In this case, the lack of familiar lateral visual cues in Vorbasse could have reduced the capability of pilots to perceive such differences. Moreover, evaluating the interaction between factors phase and difficulty (see Figure 3), it can be seen that in the easy manoeuvres the speed during the approach was higher than the one forecasted (underestimation with a mean = −0·714 knots), while during the difficult manoeuvres the speed in the same phase was lower than the estimated one (overestimation with a mean = 0·191 knots). The difference between the types of vessels employed for the manoeuvres could have determined the difference in the speed management during the approach. In the easy manoeuvres a controllable pitch propeller tanker was used. Since in this type of propulsion the shaft never stops its rotational movement, it induces a rotation to the heading of the vessel especially when the longitudinal thrust is stopped (stern transversal thrust effect enhanced when setting the propeller pitch to zero). Therefore pilots had to maintain a higher speed than forecasted in order to counteract this effect through active use of propeller thrust on the rudder. This active use of propeller thrust, on average, did not allow the expected reduction of speed to satisfy the original plan. In addition the current was coming from the stern of the vessel in that phase, helping to increase the speed over the ground. In the more difficult manoeuvres an alternate explanation for the observed lower speed than forecasted could be found in the reduced under keel clearance. Such reduced under keel clearance (down to 1·5 metres with a draft of 14 metres), enhanced the dragging effect of the two knots of current coming in that phase from the bow (possibly more than pilots expected). Moreover, even if there was no significant difference between the rest of the phases, it is interesting to note that in the more difficult manoeuvres a slight underestimation of the speed is present during the swing and the closing. A further explanation for this may be found in the action of the two knots of current interacting more significantly than pilots expected.

4.3. EngEst – Engine Power Estimation. Considering the competency of pilots to forecast the use of the main engine power (variable EngEst), a significant difference was apparent only for the factor phase. The closing phase shows a significant difference compared to the other two phases (see Figure 4). Pilots accounted in their plans for a higher use of the main engine during the closing phase (mean = 0·10). In the other two phases (approach mean = −0·064 and swing mean = −0·048), the planning estimation was slightly lower than the actual use. Moreover, a marginal interaction (Sig = 0·061) between the factors phase and difficulty was encountered (compare Figure 4(a) with 4(b), abscissa from 2 to 3). In the swing phase, pilots planned a higher need of engine than the actual use in the easier manoeuvres (mean = 0·035) but a lower need of the engine for the more difficult manoeuvres (mean = −0·131).

It is worth reiterating that these numbers are percentages. This means that the value – 0·131 expresses a difference between planned and effective use of the main engine of −13·1%. This value represents an average calculated for the entire duration of the phase. This value alone, being a difference, would not be able to define the level of power at which the main engine was working (−13% could be the result of 37% planned minus 50% effective as well as 87% planned minus 100% effective). A critical underestimation could happen for example when the power effectively required could already be close to the engine's working limits. To better explain this consideration, we can refer to the graphs obtained from manoeuvre B2 in Figure 7.

In Figure 7(a) three functions are reported. The continuous bold line shows the mean of all the pilots' EngEst scores. The dotted lines represent the upper and lower limits of the standard error of the mean with a probability of 95%. Such error was calculated using the standard deviation and considering ten subjects (see Figure 7(c)). The averaged planned and recorded engine power are reported in Figure 7(b), as fractions of 1, where 1 means 100% of available power. EngEst, reported in Figure 7(a) can be seen as the difference between those two curves graphed in Figure 7(b). Considering abscissa values from 2·5 to 3 (second half of the swing phase), in Figure 7(a) and 7(b), it can be seen how pilots expected to use the engine much less than was experienced in the simulation. The difference between the planned and the effective use reached values of 50% when the engine was working already up to 80% of its maximum power. Pilots' plans did not consider they would use the main engine that much, nor so close to its maximum availability. This may suggest that the manoeuvre could have required a different approach in that particular section to increase safety margins. This example and analysis of results not only improves understanding of shiphandling but can help pilot companies to better identify critical sections, allowing the development of more effective and safer techniques.

4.4. ThrEst – Bow Thruster Power Estimation. No significant results were found on performing an ANOVA on ThrEst. The absence of significant results in the ANOVA suggested that pilots showed a limited difference between plan and effective use of bow thruster, as confirmed also by the correlations reported in Table 4. Both B1 ( Figure 5(a)) and V1 ( Figure 5(b)) reported a significant correlation between plan and execution, confirming that pilots were able to follow their plans. The correlation between the variable ThrEst across the two easy manoeuvres was considered. The aim was to evaluate if the two manoeuvres showed similarities in the way pilots per-formed. Results confirmed pilots showed a similar performance in the two manoeuvres (r = 0·574; Sig = 0·00). This outcome supports the conclusion that the two manoeuvres, even if carried out in different ports, were essentially similar in the use of the bow thruster, showing underestimation or overestimation consistently in the same sections of these manoeuvres. This is another result that might be exploited by pilot companies to better direct the development or training activities associated with new manoeuvres.

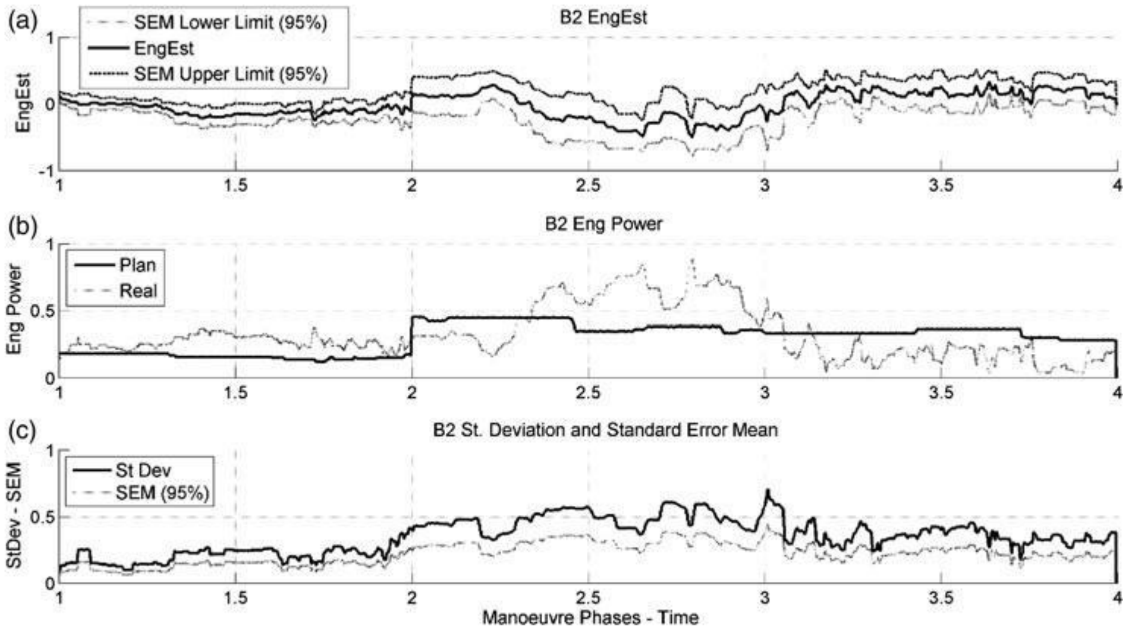



Figure 7. Manoeuvre B2 - Detailed analysis of the engine estimation.

4.5. TugEst – Tug Force Estimation. Pilots were free to decide the number of tugs that they wanted to use and their initial position. Pilots also had discretionary control over the position of the tugs during the execution. Only Tug1Est (the difference between the force expected as stated in the plan and the force effectively developed by Tug 1 during the manoeuvre) reported a significant main effect (Sig = 0·001) on the factor phase. The closing phase (see Figure 6(a), abscissa from 3 to 4) was significantly different from the other two (closing and approach (Sig = 0·06), closing and swing (Sig = 0·01)). In this case the plans prepared by the pilots forecasted a higher use of Tug 1 compared to data recorded during the simulations. The lack of other significant results in the ANOVA again suggested a general matching between plan and execution that was subsequently confirmed by the analysis of Pearson correlation coefficients. Tug 1 ( Figure 6(a), abscissa from 3 to 4) shows a clear overestimation of the bollard pull needed in the closing phase. With the help of graphs and as shown by results, it is also possible to observe that there is a generally sensitive fit between plan and execution for all three tugs. Specifically referring to Tug 2 and Tug 3, Pearson coefficients reported a lower (even if significant) correlation in manoeuvre B2 than manoeuvre V2. It should be reiterated that in manoeuvre V2, pilots experienced three crashes. This might have helped higher correlations, since data remaining was only coming from pilots that adopted a more efficient strategy and successfully completed the manoeuvre. Similarly to findings associated with the variable ThrEst, it was considered the correlation between the variable TugEst (one for each tug) measured across the two manoeuvres B2 and V2. As shown in Table 4, significant correlations were obtained. These correlations may numerically support how the strategies adopted in the use of tugs were similar in the two manoeuvres.

4.6. Study Limitations. The number of participants could represent a limitation of this study. Nevertheless, pilots spent an average of eight hours in the simulator performing these tasks, allowing a deep and detailed data collection. We recognise the value of larger data sets, and suggest that increasing the number of participants in future studies would provide more definitive results in specific manoeuvres. We ac-knowledge also the difficulties related to the somewhat unusual task that required pilots to unpack their manoeuvring mental model in a more quantifiable form represented by the Detailed Manoeuvre Plan.

5. CONCLUSIONS. In this paper, an analytical approach, comparing pilots' planned and simulated ship manoeuvres was introduced in order to more deeply understand the participant's mental models. A group of ten proficient marine pilots participated in the study. For the purposes of this paper, several variables were defined. Our expectation was that proficient pilots would have been able to provide plans that had a high degree of consistency with execution. Our aim was also to objectively quantify this matching in order to develop a methodology that could be profit-ably applied in other future comparative studies. Results obtained in the performance variables defined in Section 2.2, overall confirmed this expectation: pilots were generally able to perform according to their plans, showing only a limited number of differences in the scores recorded in the different ports, at different levels of difficulty and during different phases of the manoeuvres. Pearson correlation coefficients calculated between plans and execution also supported the expectation. Correlation coefficients between manoeuvres with the same level of difficulty further showed consistency in the way those exercises were designed, hence approached and performed. Significant differences instead pointed our attention to possible areas of improvement where pilots' approach to the manoeuvres could be discussed, reconsidered and modified. Research results confirm forecasting vessel's position was significantly more difficult for pilots during the swing than during other phases. Additional elements could have influenced these outcomes such as speed management, influenced by a different vessel's propulsion type in the easy manoeuvres and the interaction with the current in the more difficult ones. Data analysis also evidenced another marginally significant effect in the swing phase related to the estimation of the engine power. Pilots showed a tendency to slightly underestimate the use of the main engine during the most difficult manoeuvres. This was possibly related to the need in those manoeuvres to immediately respond and undertake effective actions to keep the vessel in a safe position during the rotation within a relatively smaller basin. This is just an example of how exploiting the results provided by the methodology introduced in this paper, it was possible to better analyse and unpack the complexity of shiphandling dynamics.

Ship manoeuvring requires an understanding and manipulation of complex interactions of masses and forces. It is rare that the effects of these interactions observe linear laws. This makes their appreciation and prediction a considerable task, especially when carried out without the support of appropriate tools and training. This very fact has led other researchers to explore the possibility, through fast time simulations, of making more accurate real time predictions of a vessel's behaviour while manoeuvring (Benedict, 2012). Nevertheless, the seamless integration of operators and state of the art technology (when available on ships' bridges) continues to evolve. Pilots perform their job with different types of vessels, each of them with its unique configuration of bridge equipment and personnel. They have to quickly adapt to the situation, making critical judgements as to the feasibility and the safety of the manoeuvre that they will immediately execute. In this study it was shown and quantified how such judgement could be sensitive to inaccuracies. Those inaccuracies may become more relevant as the situation departs from relatively stable and more linear conditions, as highlighted by the results obtained in the swing phase.

5.1. Future Applications and Added Value. The method described in this paper, if systematically adopted, provides a valid and reliable basis to better develop training and test manoeuvring techniques. Analysing results with this methodology could help to clearly identify optimal ranges of distances, speeds or use of available means, thus allowing the development of safer and more efficient manoeuvres. Remembering that the comparisons and the results obtained are based on simulated results, this research argues that it will be of the utmost importance in the future to apply the same methodology to real life shiphandling contexts. Using systematic feed-back from similar manoeuvres in real situations, it will be possible to refine reliability and further validate simulated models.

Portable Pilotage Units (PPUs) and ships' Voyage Data Recorders (VDRs), engine logs, video and audio recordings can be exploited in order to collect this data in a work-place context. Within this naturalistic approach, it will not be possible to decide a priori the level of difficulty of the berthings so performed. Mooring operations, once recorded, can be grouped in different levels, comparing the conditions encountered, ac-cording for example to a "level matrix" similar to the one used for the simulator assessment here introduced. Such an approach may open the opportunity for new avenues of research and provide applications that may include: (a) the creation of standardised simulated exercises to select, train, evaluate and certify pilots based on national standards; (b) identification of more realistic construction criteria for actual/future port developments; (c) more reliable port operations safety criteria through more accurate risk assessments.

Based on the findings and the methodological approach reported in this initial foundation research, further empirical analysis on data differently sourced needs to be carried out. Comparative studies with different groups of shiphandlers at different levels of experience and engaged in different manoeuvres, used as models, would help to standardise scales able to better define the dimensions of shiphandling expertise.

ACKNOWLEDGEMENTS

We would like to thank the Australasian Marine Pilots Institute and Smartship Simulator for the priceless support offered in the realization of this work. We are deeply grateful to the pilots that patiently bore with us and with the intervention of too many obnoxious devices. A special acknowledgement goes to Dr. Irene Penesis and Dr. Elkana Ngwenya for their patient guidance through the minefields of statistical analysis.

# APPENDIX

Table A1. ANOVA Results for XTD.

| Source | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Model | 200458·903[a] | 12 | 16704·909 | 16·462 | ·000 |
| Diff | 288·904 | 1 | 288·904 | ·285 | ·595 |
| Port | 699·833 | 1 | 699·833 | ·690 | ·408 |
| Phase | 16711·269 | 2 | 8355·634 | 8·234 | ·000 |
| Diff * Port | 1174·076 | 1 | 1174·076 | 1·157 | ·285 |
| Diff * Phase | 3756·822 | 2 | 1878·411 | 1·851 | ·162 |
| Port * Phase | 139·416 | 2 | 69·708 | ·069 | ·934 |
| Diff * Port * Phase | 1313·119 | 2 | 656·560 | ·647 | ·526 |
| Error | 101472·775 | 100 | 1014·728 | | |
| Total | 301931·677 | 112 | | | |

a. R Squared = ·664 (Adjusted R Squared = ·624)

Table A2. ANOVA Results for SpdEst.

| Source | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 13·088[a] | 11 | 1·190 | 4·855 | ·000 |
| Intercept | 2·957 | 1 | 2·957 | 12·067 | ·001 |
| Diff | ·167 | 1 | ·167 | ·683 | ·410 |
| Port | 1·015 | 1 | 1·015 | 4·141 | ·044 |
| Phase | ·749 | 2 | ·375 | 1·529 | ·222 |
| Diff * Port | ·017 | 1 | ·017 | ·067 | ·796 |
| Diff * Phase | 9·981 | 2 | 4·991 | 20·364 | ·000 |
| Port * Phase | ·671 | 2 | ·335 | 1·369 | ·259 |
| Diff * Port * Phase | ·014 | 2 | ·007 | ·029 | ·972 |
| Error | 24·506 | 100 | ·245 | | |
| Total | 40·358 | 112 | | | |
| Corrected Total | 37·594 | 111 | | | |

a. R Squared = ·348 (Adjusted R Squared = ·276)

Table A3. ANOVA Results for EngEst.

| Source | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Model | ·951[a] | 12 | ·079 | 2·189 | ·018 |
| Diff | ·061 | 1 | ·061 | 1·682 | ·198 |
| Port | ·000 | 1 | ·000 | ·011 | ·918 |
| Phase | ·600 | 2 | ·300 | 8·293 | ·000 |
| Diff * Port | ·008 | 1 | ·008 | ·217 | ·643 |
| Diff * Phase | ·208 | 2 | ·104 | 2·876 | ·061 |
| Port * Phase | ·042 | 2 | ·021 | ·586 | ·558 |
| Diff * Port * Phase | ·054 | 2 | ·027 | ·745 | ·477 |
| Error | 3·619 | 100 | ·036 | | |
| Total | 4·570 | 112 | | | |

a. R Squared = ·208 (Adjusted R Squared = ·113)

Table A4. ANOVA Results for ThrEst.

| Source | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | ·407[a] | 5 | ·081 | 1·023 | ·413 |
| Intercept | ·206 | 1 | ·206 | 2·596 | ·113 |
| Diff | ·000 | 0 | – | – | – |
| Port | ·000 | 1 | ·000 | ·002 | ·960 |
| Phase | ·293 | 2 | ·146 | 1·842 | ·168 |
| Diff * Port | ·000 | 0 | – | – | – |
| Diff * Phase | ·000 | 0 | – | – | – |
| Port * Phase | ·114 | 2 | ·057 | ·714 | ·494 |
| Diff * Port * Phase | ·000 | 0 | – | – | – |
| Error | 4·294 | 54 | ·080 | | |
| Total | 4·907 | 60 | | | |
| Corrected Total | 4·700 | 59 | | | |

a. R Squared = ·087 (Adjusted R Squared = ·002)

Table A5. ANOVA Results for Tug1Est.

| Source | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 1·204[a] | 5 | ·241 | 3·315 | ·012 |
| Intercept | ·000 | 1 | ·000 | ·002 | ·964 |
| Diff | ·000 | 0 | – | – | – |
| Port | ·016 | 1 | ·016 | ·224 | ·638 |
| Phase | 1·183 | 2 | ·592 | 8·143 | ·001 |
| Diff * Port | ·000 | 0 | – | – | – |
| Diff * Phase | ·000 | 0 | – | – | – |
| Port * Phase | ·032 | 2 | ·016 | ·218 | ·805 |
| Diff * Port * Phase | ·000 | 0 | – | – | – |
| Error | 3·342 | 46 | ·073 | | |
| Total | 4·558 | 52 | | | |
| Corrected Total | 4·546 | 51 | | | |

a. R Squared = ·265 (Adjusted R Squared = ·185)

Table A6. ANOVA Results for Tug2Est.

| Source | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | ·210[a] | 5 | ·042 | ·674 | ·645 |
| Intercept | ·014 | 1 | ·014 | ·219 | ·642 |
| Diff | ·000 | 0 | – | – | – |
| Port | ·006 | 1 | ·006 | ·101 | ·752 |
| Phase | ·196 | 2 | ·098 | 1·572 | ·219 |
| Diff * Port | ·000 | 0 | – | – | – |
| Diff * Phase | ·000 | 0 | – | – | – |
| Port * Phase | ·014 | 2 | ·007 | ·114 | ·893 |
| Diff * Port * Phase | ·000 | 0 | – | – | – |
| Error | 2·871 | 46 | ·062 | | |
| Total | 3·090 | 52 | | | |
| Corrected Total | 3·081 | 51 | | | |

a. R Squared = ·068 (Adjusted R Squared = −·033)

Table A7. ANOVA Results for Tug3Est.

| Source | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | ·276[a] | 5 | ·055 | ·719 | ·613 |
| Intercept | ·017 | 1 | ·017 | ·226 | ·637 |
| Diff | ·000 | 0 | – | – | – |
| Port | ·002 | 1 | ·002 | ·030 | ·863 |
| Phase | ·244 | 2 | ·122 | 1·591 | ·215 |
| Diff * Port | ·000 | 0 | – | – | – |
| Diff * Phase | ·000 | 0 | – | – | – |
| Port * Phase | ·030 | 2 | ·015 | ·198 | ·821 |
| Diff * Port * Phase | ·000 | 0 | – | – | – |
| Error | 3·529 | 46 | ·077 | | |
| Total | 3·842 | 52 | | | |
| Corrected Total | 3·804 | 51 | | | |

a. R Squared = ·072 (Adjusted R Squared = −·028)

## 6.3. Paper III

# Measuring mental workload and physiological reactions in marine pilots: building bridges towards redlines of performance.

Luca Orlandi and Benjamin Brooks

*(University of Tasmania, Australian Maritime College, National Centre for Ports and Shipping, Launceston, Tasmania 7250, Australia)*

(E-mail: luca.orlandi@utas.edu.au; Tel: +61 3 6324 3999)

**Abstract**

This paper investigates the effects of shiphandling manoeuvres on mental workload and physiological reactions in ten marine pilots. Each pilot performed four berthings in a ship simulator. Those berthings were differentiated by two factors, level of difficulty and familiarity with the port. Each berthing could also be divided into five phases, three during the execution and two resting periods, one before and one after the execution (dedicated to baseline physiological data collection). Mental workload was measured through two self assessment scales: the NASA TLX and a Likert scale. Power spectral densities on Beta bands 1 and 2 were obtained from EEG. Heart rate and heart rate variability were obtained from ECG. Pupil dilation was obtained from eye tracking. Workload levels were higher as berthings increased in difficulty level and/or the pilots completed the berthings in unfamiliar ports. Responses differed across specific phases of the berthings. Physiological responses could indirectly monitor levels of mental workload, and could be adopted in future applications to evaluate training improvements and performance. This study provides an example of an applied methodology aiming to define an upper redline of task demands in the context of marine pilotage.

**Practitioner Summary**

Mental workload and physiological reactions were investigated in marine pilots during simulations. As expected, correlations between workload and physiological variables were identified. Workload peaked during one phase of the berthing. Application of this method with a larger sample may provide clearer indications of a "redline" of workload in this occupational group.

**Keywords**

1. Mental workload 2. Marine Pilotage 3. Physiological measures.

## 1. Introduction

Shipping represents the major player in transportation with commercial vessels carrying around 90% of the world trade. As reported by the International Chamber of Shipping, the maritime industry generates an annual income of over half a trillion US dollars in freight rates, with a worldwide population of seafarers serving on internationally trading merchant ships on the order of 466,000 officers and 721,000 ratings (ICS, 2016). Even though some authors consider the shipping industry having a fairly good safety record (Hetherington et al., 2006), it does not compare particularly well to other mass transport modes, and is not necessarily improving its performance.

A study conducted in the US comparing different transport modalities, reports that the workplace fatality rate per 1000 employees in the maritime transportation (0.24) is four times as high as the one in the air transportation (0.06) (Savage, 2013). In 2013, a report from the IMO (international Maritime Organization) Correspondence Group on E-navigation provided some statistics based on the IHS Fairplay casualty database (considered the most complete and reliable maritime data source in the world). This report highlighted how the total number of navigational accidents on cargo, passengers and offshore ships increased between 2001 and 2010 (from less than 400 in 2001 to more than 700 in 2010). The report showed also how the number of accidents per ship increased from 0.5% in 2001 to 1% in 2010. Of the total number of accidents considered, 22% were groundings, 22% were collisions and the rest were classified as other types (IMO, 2013).

Many systemic factors have been identified as contributing to maritime accidents (Perrow, 2011), such as the social organization of the personnel on board, economic pressure, 'hidden' ownership structures, and difficulties in international regulation. At an individual level, long contracts, limited sleep opportunities between shifts and short turn-around times can create fatigue, stress and work pressure (McNamara et al., 2000). An example can be found in a Scandinavian study that compared the psychosocial working conditions and mental health of a group of maritime engine officers with a group of British shore based professional engineers. The study highlighted that while the British shore based engineers reported significantly higher role ambiguity the Swedish engine officers perceived a significantly higher degree of role conflict and higher perceived stress (Rydstedt & Lundh, 2010). A Canadian report (CMPA, 2017) explained how one of the most effective measures that are adopted in the shipping industry to mitigate groundings and collisions is the use of Marine Pilots. The reports highlighted how piloted ships are able have their risk reduced 44 times compared to not piloted ships (from 0.094 to 0.0021 probability of accident per vessel). The risk of collision and grounding drops 12 times more if a piloted vessel has also tugs in assistance (from 0.0021 to 0.00018 probability of accident per vessel).

Within this context, marine pilots were chosen as participants in this study. Marine pilots are ship's captains that are specifically trained and certified to manoeuvre vessels within critical coastal and port waters. They embark a ship outside port waters and then work with the bridge team to navigate the ship to berth. While ship's Captains still retain the full charge of the vessel, pilots generally take the "conduct". They manoeuvre the ship in enclosed and or critical waters until a safer position is reached or the vessel is alongside the assigned mooring. Piloting involves a complex interaction between the pilot and a bridge team, tug masters, a vessel traffic service and electronic equipment.

Pilots can be defined as "experts" following the definition of those who acquired noticeable skills or knowledge of a particular subject, through training and practical experience, capable of recalling complex, task specific patterns gaining access to the right information (Scardamalia & Bereiter, 1991). As experts, pilots are expected to be specialists having specialised knowledge (Mieg, 2001); they are able to restructure, reorganize, and refine their representation of knowledge, applying it more efficiently into their environment. Pilots, with their expertise being the result of a complex adaptations of mind and body, should be able to exploit substantial self-monitoring and control mechanisms to the tasks and goals imposed to them by the environment (K. A. Ericsson & Lehmann, 1996). Their actions should be smoother and more efficient, and performance should be achieved with a minimal effort, running essentially automatically, with minimal cognitive control (Posner & Snyder, 2004). They should be able to run more processes in parallel, thanks to the reduction in the mental workload due to automaticity (Shiffrin & Schneider, 1977).

Mental workload is a multidisciplinary concept (M. S. Young et al., 2015) and has long been recognized as an important element of human performance (Eggemeier et al., 1991; Parasuraman et al., 2008), particularly important in high risk environments (Jou et al., 2009) and those demanding high levels of reliability (Carswell et al., 2005; Yurko et al., 2010). Mental workload varies around a combination of task demands and resources that a particular individual has available (Noyes et al., 2004; M. Young & Stanton, 2005). From this "resource-based view", mental workload can be seen as the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support and experience. For the purposes of this study, mental workload followed the definition of subjects' direct estimate or comparative judgment of mental or cognitive effort experienced at a given moment (Luximon & Goonetilleke, 2001).

Even though qualitative studies have been conducted (M. H. Lützhöft & Nyce, 2006), we still cannot define what would be an acceptable level of workload and what are the physiological implications for pilot's workplace health and safety. The levels of mental workload in shipping, compared to other areas of the transport industry are relatively unknown, as indicated in a recent review (M. S. Young et al., 2015), with only few papers published in the last thirty years. A 2008 study, involving 20 Norwegian Navy cadets, investigated the relationship between workload, navigational method and performance in a shipping simulator. The use of electronic chart and information systems (ECDIS) was compared against traditional methods of navigation (paper charts). The use of ECDIS highlighted advantages in terms of ship position accuracy and handling, but did not provide significant differences in workload, as measured by heart rate variability and skin conductance (Gould et al., 2009). A previous study used heart rate variability to assess the workload of a single officer of the watch. Significant differences in workload were found while conducting a real vessel in six different geographical areas (Murai et al., 2004).

Anecdotally, marine pilots would appear to face considerable variation in workloads when managing different manoeuvres, working in changing environmental conditions and due to the dynamic nature of commercial shipping. For pilotage organisations, the main concern is that workload experienced might breach acceptable levels, exceeding what has been defined as the "red line" of workload/performance (Brookhuis, Waard, & Fairclough, 2003). Given that a single accident has the potential to close an entire port, establishing this "redline" is of value to port authorities and pilotage organisations. In a study conducted on car drivers, Horrey and Wickens introduced the possibility of analytically calculating the impact of competing pairs of tasks on workload and performance (Horrey & Wickens, 2003) and were able to account for almost 100% of the variance in task performance and hazard response.

With these elements in mind, this study aims to quantify and evaluate the impact on pilots workload of different shiphandling conditions while berthing ships in a simulator, adopting concurrent self reported and physiological measures. The hypotheses investigated are:

- Berthings with different levels of difficulty should elicit different levels of self reported scores as well as different levels of physiological reactions.
- Berthings performed in a foreign port should elicit higher levels of workload.
- Concurrent measurements known from the literature to be related to workload, should show similar trends.

## 2. Methodology

### *2.1 Experimental Design*

To investigate pilots' workload and its related measures, four different berthing manoeuvres were set as experimental conditions. Exactly the same four berthings were conducted by each participant, even though in random order, to mitigate a possible learning effect. Each berthing included the whole process necessary to transfer the ship from a defined initial position to a berth within constrained port waters, with the use of own and/or external means of propulsion (i.e. tug boats to assist, when allowed). The berthings were presented to pilots before being performed in the simulator, since every participant was required to provide a plan, such as the one normally discussed by pilots and ship masters before a ship enters into a port (Wild & Constable, 2013).

### *2.2 Participants*

The participants to this study were a group of ten marine pilots from an Australian pilot company. They were all males in good health, as required by national professional medical standards set by the Australian Maritime Safety Authority (AMSA, 2010). An Analysis of Variance (ANOVA) for age and service confirmed no significant difference between the participants and the rest of the pilot population working for the same company. All the pilots involved in the research had more than ten years of previous experience in pilotage, even if not in the same Company. The number of participants is comparable to similar studies focused on niche professional categories (Di Stasi et al., 2015; Itoh et al., 1990; Sirevaag, Kramer, Reisweber, Strayer, & Grenell, 1993). Before completing the berthings in the simulator, pilots had one (or more, if required) face to face session(s) with the researcher in order to provide their passage plans, a detailed descriptions of their shiphandling expectations sketched on a navigational chart (Luca Orlandi et al., 2015). Once the passage plans were completed for all the four required berthings, each pilot spent a whole day at the simulator facility to perform the exercises (five in total, including the familiarization). The two simpler berthings had a duration of about 1 hour, while 2 hours were necessary to complete the two most difficult ones. During the berthings (and also before and after, for specific physiological measurements) the studied variables were continuously recorded, obtaining for each pilot between 6 and 8 hours of continuous physiological data collection, as elicited by the different manoeuvring scenarios. The authors assert that all procedures contributing to this work comply with the ethical standards of our University and the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

### *2.3 Independent Variables*

In each of those berthings, three main factors were controlled. These three main factors were: (a) “port familiarity” (from now on referred as “port”), (b) “difficulty”, and (c) “phase”.

#### *2.3.1 Port*

The first factor, “port”, took into account whether the berthings were conducted in the participant pilot’s homeport (the port where they were regularly working) or in a foreign port. The foreign port was a virtual port only present in the simulator software. This port was chosen to avoid any possibility of a learning effect associated with previous manoeuvring experience the subjects may have possessed and to provide support for methodology reliability. The pilots’ homeport in the tables and graphs presented is coded “0”, while the virtual port is coded “1”.

#### *2.3.2 Difficulty*

The second factor was the ship-handling level of “difficulty”. To control the level of difficulty, specific berthings parameters were altered as summarised in table 1. The easy level is coded “0” while the difficulty level is coded “1”.

**Table 1. Levels of Difficulty – Adopted in both Ports.**

| | Pier - Spatial constraints | Environmental conditions and forces | Vessel characteristics | Tugs | Interactions with traffic | VTS Comms [1] |
|---|---|---|---|---|---|---|
| **Level 0**<br><br>**Easy** | Big Swing Basin<br><br>(3 times Vessel LOA [4]) | Current: 0.7 Knt<br><br>Wind: 15 Knt<br><br>Good Visibility | Single Controllable Pitch Propeller [2]<br><br>Bow Thruster [3] | None | 1 Interacting but not Interfering vessel | General Info<br><br>No reporting Points |
| **Level 1**<br><br>**Difficult** | Small Swing Basin [5]<br><br>(1,5 times Vessel LOA) | Current: 2 Knt<br><br>Wind: 25 Knt<br><br>Poor to no Visibility - Heavy Rain | Single Fixed Pitch Propeller<br><br>No Thrusters | As required by Pilot | 1 Interacting<br>1 Interfering vessels | General Info and Traffic Advice<br><br>Reporting Points |
| **Notes** | (1) Vessel Traffic Management station present in a port and managing ships via radio communications;<br>(2) Propeller capable to change the water thrust direction changing the angle of the blades instead of direction of rotation;<br>(3) A thruster is a propeller positioned perpendicular to the ship keel axis. Placed on the bow or on the stern, induces transversal / angular motion;<br>(4) Length Over All, maximum length of a vessel;<br>(5) Wider area, within constrained waters, where ships have sufficient room to rotate and revert their direction. | | | | | |

Level 0 reproduced a comparable level of difficulty of routine operations. Level 1 aimed to engage pilots with a scenario slightly exceeding the safety limits established in the pilots' homeport, without losing construct validity. Note that level 1 of difficulty implied that the safety limits currently adopted in pilots' homeport were exceeded. This experimental condition was introduced to explore results in a situation that even though not present at the beginning, could develop during a real berthing. Even if ports adopt specific environmental criteria to set the safety limits beyond which berthings cannot be performed, it can happen that these limits could be reached and exceeded while a berthing is in progress and cannot, at that point, be interrupted.

*2.3.3 Phase*

Each berthing required pilots to complete a mooring using the side of the ship opposite to the berth position on commencement of the exercise. This implied that for each berthing the ship had to swing (rotate 180°) before she could be moored. Each berthing therefore had to be developed through three main phases.

These "phases" provided the additional factor for the analysis (see figure 1). The three phases were: (1) the "approach" (from the initial position until the start of the swing), (2) the "swing" (from the start of the swing until the rotation was completed and stabilised), and (3) the "closing" (from the end of the swing until a defined distance from the berth). Through the analysis of the video recordings and the simulator replays, it was possible to identify the instant in time were each berthing was transitioning between phases. To provide an example, it was used as the instant when the Swing started, the clear order or indication given by the pilot that he was committing to the vessel rotation. This was also cross checked with the indication that the ROT was increasing and subsequent actions with tugs and ship propulsion were consistent with what was required to induce the swing (rudder angle to one side, tugs lifting off and / or pushing at the beam of the vessel, etc..). The start of the Closing was identified not only through pilots' direct indications, but also cross checking when the ROT was minimised again (swing completed and vessel stabilised on a constant heading). Other elements were considered as well, such as actions with ship propulsion and tugs, consistent with the end of the rotation and closing to the berth. In the graphs and tables presented, the phases are coded: “1” for the approach, “2” for the swing and “3” for the closing.

Two additional phases were also introduced, specifically for the collection of physiological variables. The “Baseline Pre” or phase “0” was a period of data collection of circa 5 minutes, immediately before the execution of the berthings. The “Baseline Post” or Phase “4” was a period of physiological data collection immediately after the execution of the berthings. Participants were required to stand still, in the same position where they mainly remained during the execution of the berthings, to allow a data collection during a period of time when they were not involved in any activity related to the exercises. These recordings were meant to provide physiological data at rest, before and after the execution, for comparison purposes with the data obtained during the three active phases of the berthing.

### *2.4 Dependent Variables*

The dependent variables herein listed, were collected with the aim to quantify directly (with NASA TLX and a purposely built self-assessment scale) and indirectly (as a response measured by physiological variables), pilots’ mental workload while they were performing the different berthings (Luca Orlandi, Brooks, & Bowles, 2014).

#### *2.4.1 NASA TLX*

Subjective mental workload was assessed through the adoption of two self-assessment scales. The first of these was the National Aeronautics and Space Administration-Task Load Index (NASA-TLX). NASA-TLX (Hart & Staveland, 1988) is a multidimensional scale for which the overall mental workload is a function of 6 subscales: 1. Mental Demand (MD), 2. Physical Demand (PD), 3. Temporal Demand (TD), 4. Own Performance (OP), 5. Effort (EF), 6. Frustration Level (FR). At the end of each berthing a NASA TLX form, in a Microsoft® Excel® electronic format, was completed by the shiphandlers. In a dedicated sheet for each berthing, the Excel workbook presented six sliding rulers (with 21 possible positions). Using those sliding rulers, pilots were able to select the experienced raw score in any subscale. The initially required weighing procedure of the scales was performed only once and at the end of the first berthing. Using the results obtained from the initial weighting procedure and combining them with the raw scores recorded by the sliding rulers, it was possible to obtain, for each run, the weighted scores in the 6 sub scales and the total NASA TLX score. These results (weighted scores), collected at the end of each run, were used for the following data analysis.

#### *2.4.2 Self Assessment Likert Scale*

The second self assessment workload measurement was obtained through the analysis of audio recordings, captured during the entire execution of the four berthings. Before the berthings were started, shiphandlers were instructed about the use of a Likert scale built specifically for this study, to verbally report their level of “involvement or workload”. The scale (always kept in shiphandlers” sight), reported 7 different levels of “exercise difficulty” (see Appendix A), meaning the personal level of workload experienced or effort necessary, in order to be able to manage the situation at the time of the question. The shiphandler, during the execution of the berthing, was briefly reminded (every two minutes circa), with a quick question asked by the researcher (i.e. “How do you feel?”), to simply report a number according to the scale above described. The results so obtained, once time stamped, were then transformed into a continuous numerical function.

The physiological variables chosen for this study, are known from the literature to be associated with workload, as well summarised in a recent review focused on car drivers and aircraft pilots (Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014):

#### *2.4.3 ECG – HR and LF/HF*

Heart Rate (HR) and Heart Rate Variability (HRV) have been extensively exploited to study the association between psychological processes and physiological reactions (Berntson, 1997;

De Waard, 1996). Higher workload is associated both with higher HRs and with higher values in the LF/HF function (Van Amelsvoort et al., 2000).
A Smartex® Wearable Wellness System® was used (www.smartex.it) for the recording of the Electrocardiogram signal. This system, uses two electrodes inbuilt in the fabric of the wearable t-shirt. The input sensitivity is ±5 mV. The sampling rate is 250 samples per second, covering a bandwidth from 0.05 Hz to 30 Hz with a 12 bits resolution. The collected ECG signals were processed using the free software Kubios HRV . Kubios software was used to visually inspect the raw ECG recordings and clean artefacts using the provided filtering functions. Being able to identify the instant of each heart beat on a time line, allows to obtain two continuous functions (one the inverse of the other) in the domain of time:

- The heart rate function (HR), which reports the number of beats per minute at any instant in time;
- The Inter beat Interval (IBI) function, which reports the period (in milliseconds) recorded between each two consecutive heart beats, at any instant in time.

The HR function, once normalised (see dedicated paragraph 2.6), was used in further analysis to obtain the HR dependent variable.

The IBI function was further processed to obtain the LF/HF dependent variable. The LF/HF, is the ratio between the integral of Power Spectral Density (PSD) calculated in the low frequencies band (LF) and the Power Spectral Density (PSD) calculated in the high frequencies band (HF). The PSD functions are obtained from the spectral analysis of the IBI function. The spectral analysis of the IBI function was performed using the algorithms provided in the open source software HRVAS (Ramshur, 2010). HRVAS software is a graphical interface developed to implement functions built in Matlab scripting code (MathWorks, 2013) for ECG analysis. Those Matlab functions were extracted and re-integrated in newly developed scripts, able to batch process all the ECG recordings collected during the berthings. An ectopic detection algorithm (still provided in the HRVAS software) was initially used to exclude those values in the IBI functions that were exceeding 20% of previous sample value or exceeding 3 times the standard deviation of the signal or using a median filter where the tau value was set to 4. The ectopic values so identified were substituted using a “spline” interpolation. The cleaned IBI signals were de-trended. To calculate the LF/HF values, it was chosen a shifting window of 128 seconds. To obtain the continuous LF/HF function in the domain of time, all the calculations were reiterated shifting the 128 seconds window, of 1 second at a time. This process implied that two consecutive windows would have overlapped for 127 seconds. Reiterating the spectral analysis for each sequential 128 seconds interval, allowed to obtain, for each IBI recording, a continuous LF/HF function in the domain of time. The LH/HF values so obtained were normalised, following the procedure explained in the dedicated paragraph 2.6.

*2.4.4 Pupil Dilation*

Pupil dilation is also known from the literature as an indirect measurement of mental workload (Batmaz & Öztürk, 2008; Iqbal, Adamczyk, Zheng, & Bailey, 2005). Pupil Dilation was recorded using a pair of light-weight eye tracking goggles ASL® Mobile Eye XG® (www.asleyetracking.com). This eye-tracker system comprises two cameras each sampling at 30 Hz: a forward facing scene HD camera and an eye camera, capturing the infrared corneal reflection and pupil position of the right eye. The pupil dilation was recorded by the eye tracker device as the pupil diameter measured in number of pixels, and timestamped consistently with the video recording carried out at the same time. The diameter was collected 30 times per second. When data was lost, for temporary interruption of the pupil tracking, it was replaced using a linear interpolation between the values at the extremities of the gap. Signals were cleaned from outliers exceeding 20% of previous sample value or exceeding 3 times the standard deviation of the signal. The pupil dilation values were directly collected as a function in the domain of time for all the berthings. Phase 0 and 5 were not recorded before and after the execution of berthings, since the eye trackers were not active at that point in time.

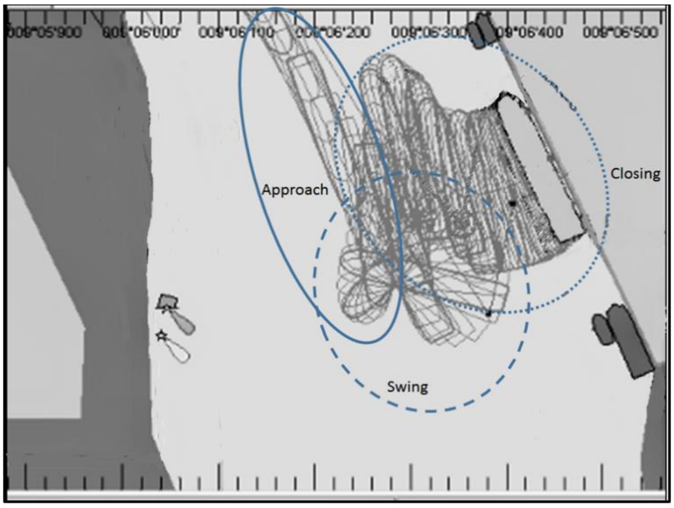
*2.4.5 EEG – Bands Beta 1 and Beta 2*

The electro encephalograph signals were recorded using an Emotiv® Epoc® wireless device (www.emotiv.com), featuring 14 channels: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF42. The sampling rate was 128 samples per second with a resolution of 14 bits (1 LSB = 0.51μV), covering a bandwidth from 0.2 Hz to 43 Hz, with digital notch filters at 50 Hz and 60 Hz. Similarly to the data processing performed with the ECG, also with EEG signals it is possible to conduct a spectral analysis. Frequency bands for the study of EEG are generally defined as follows: Delta (below 4 Hz), Theta (4 - 7 Hz), Alpha (8 – 13 Hz), Beta (13 – 36 Hz). Several studies have found an association between specific bands of the EEG and mental workload (Di Stasi et al., 2015; Dussault, Jouanin, Philippe, & Guezennec, 2005; Kohlmorgen et al., 2007). For the purposes of this study the PSD was specifically calculated in the bands Beta 1 (13 – 20 Hz) and Beta 2 (20 – 36 Hz). Higher values recorded in these bands have been associated with higher mental workload (Koester, 2003b). The values were obtained as the ratio between the integral of the PSD function calculated in the chosen band (Beta 1 or Beta 2) over the total power calculated in all the bands (obtained as the integral of the PSD function in the bands Delta + Theta + Alpha + Beta 1 + Beta 2). The Beta 1 and Beta 2 values so obtained were normalised following the procedure explained in the dedicated paragraph 2.6. The normalised values were calculated separately, for each one of the fourteen channels.

***2.5 Procedure***

Four berthings were conducted by each pilot. Two of the four berthings were set in the pilots' homeport (one berthing for each level of difficulty), two were set in the foreign port (one berthing for each level of difficulty). Figure 1 shows one screenshot taken from the simulator interface, showing one of the four berthings (easy berthing in the virtual port). It is possible to notice in light grey the outline of the vessel used (M/V Torm Laura). The empty outlines creating the shaded area represent the swept path covered by the vessel during its movement.

**Figure 1. A berthing as shown by the simulator interface.**
**Three phases highlighted by circles.**




Berthings were also coupled across the factor "port" (for the same level of difficulty); i.e. the easy berthings in the two ports (as well as the difficult berthings) were, as much as possible, kept technically similar (e.g., vessels used, distances to be covered, etc.) to promote baseline data formation on pilot performance and assure reliability of the assessment process. Spatial constraints due to port dimensions were purposely maintained similar, modifying the virtual port in order to match the homeport dimensions as summarised in table 2.

**Table 2. Proportions between vessels and port dimensions.**

| Ship | LOA (m) | Ratio between Ships | Breadth (m) | Disp (ton) |
|---|---|---|---|---|
| **Torm Laura (diff Lvl 0)** | 183 | 0.7 | 32 | 54925 |
| **Arcturus (diff Lvl 1)** | 269 | 1.45 | 48 | 143200 |
| **Ratio** | **Torm LOA** | **Torm Breadth** | **Arcturus LOA** | **Arcturus Breadth** |
| **Basin diameter (470 m)** | 2.6 | 14.7 | 1.7 | 9.8 |
| **Channel width (300 m)** | 1.6 | 9.3 | 1.1 | 6.2 |

The Maritime Safety Queensland Simulator located in Brisbane was used (Smartship® Simulator www.smartshipaustralia.com.au). This "Full Mission Bridge" simulator is classified as Class A (NAV) according to the standards issued by the classification society Det Norske Veritas Germanischer Lloyd (DNV_GL_AS, 2014). It is capable of simulating a total shipboard bridge operation, including the capability for advanced manoeuvring in restricted waterways. Before the experimental berthings, pilots were required to perform a very simple mooring with a vessel different from those used in the experimental runs. This first berthing was used as a familiarisation run to ensure participants had a standardised level of familiarity with the bridge environment and the navigation equipment available. The experimental berthings were then used in random order to record all the data. To provide realism to the berthings, during their execution, the researcher was present on the simulator bridge and he was generally acting as the ship's Master or the bridge member most suitable for the specific interaction.

While pilots were performing their berthings, physiological data was continuously recorded. In order to obtain the measurements without interfering with pilots' performance, only portable, wireless and sufficiently comfortable recording devices were chosen. All the results obtained from physiological variables, were normalized across all the berthings performed by each participant as better described in the following paragraph.

***2.6 Quantile Normalization of Results***

In order to be able to compare physiological data, relative to different measurements and several subjects, a quantile normalization of the scores was adopted (B. M. Bolstad et al., 2003). The normal distribution was selected as reference distribution for the transformation. This procedure was applied to ECG (Heart Rate and LF/HF), EEG (Bands Beta 1 and 2) and Pupil Dilation.

The empirical cumulative distribution of the raw scores was calculated in Matlab (MathWorks, 2013). All the scores (of each one of the dependent variables) recorded in the four berthings performed, were considered all together for each subject. For each raw score (considering the same subject and all the four berthings together) it was then possible to calculate its percentile value. Using the percentile value "X" it was possible to obtain the correspondent value "Y" of a normal distribution (mean = 0 and standard deviation = 1), using the inverse function of the empirical cumulative distribution. This process was reiterated for each subject and for each dependant physiological variable. This process allowed to obtain a normal distribution of all the scores recorded by any subject, for any specific physiological variable and in all the 4 four berthings. The subjects' normal scores could then be used to describe a continuous function in the domain of time that could be used in further analysis. Averages across participants for the three considered factors in the study, were then calculated using previously normalized scores.

### *2.7 Data Analysis*

The continuous measurements of the dependent physiological measurements and the self-assessment Likert Scale scores, were collected and processed for each berthing completed by each participant. The results were averaged for each berthing and for each participant, within each phase previously identified as "baseline pre", "approach", "swing", "closing" and "baseline post".

For the calculation of the EEG - Beta 1 and Beta 2 scores, an additional step was required: since the EEG recording included 14 channels, the normalised scores were averaged separately for each channel in each phase. Then the median, among the 14 averages so obtained, was chosen to represent the total Beta 1 (or Beta 2) score in each specific phase.

The dataset used for the statistical analysis presented in this study, included different number of cases according to the dependent variable considered:

- ECG (HR and LF/HF) and EEG (Beta 1 and 2) variables included the phases "Baseline Pre" and "Baseline Post", introduced to capture pilots' resting values, for comparison purposes with scores collected while manoeuvering.
  Dataset cases = 200 (10 Subjects X 4 Berthings X 5 phases).

- Pupil Dilation, was recorded only during the execution of the berthings (phases "approach", "swing" and "closing") since eye trackers were activated and were recording only during the active shiphandling. The time required to setup and calibrate the equipment between runs, did not allowed a data collection also in the "Baseline Pre" and "Baseline Post" phases.
  Dataset cases = 120 (10 Subjects X 4 Berthings X 3 phases).

- Self Assessment Likert Scale, was not recorded during baseline phases (since pilots were at rest).
  Dataset cases = 120 (10 Subjects X 4 Berthings X 3 phases).

- NASA TLX was collected at the end of each berthing, so no separate scores for the phases were available.
  Dataset cases = 40 (10 Subjects X 4 Berthings)
  For the calculation of the Pearson correlation coefficients with the other dependant variables, the total scores were repeated in the three phases (Approach, Swing and Closing) for each berthing.

In the analysis, when data was not available, it was treated as missing (See following paragraph "Missing Data").

For all the variables so collected, a 3 Way Analysis of Variance (ANOVA) was conducted (using the statistical package IBM SPSS (IBM_Corp., 2010)) on the factors "difficulty", "port" and "phase". Statistical significance was accepted at the $p < 0.05$ level. It was chosen an ANOVA "between subjects" instead of a design "within" subjects, purposely aiming for a less sensitive statistics, in consideration of the missing data that will be better explained in the next paragraph. When interactions between factors were found, further analysis was conducted using 2 and 1 Way ANOVA as necessary, using Bonferroni adjustment and maintaining in the calculations the 3 Way error term.

Using the same statistical package IBM SPSS, it was also performed a Pearson correlation analysis (Sig 2-tailed) between all the dependent variables. The statistics were obtained considering, for each pair of variables, only the cases with valid data for that pair (due to the different number of dataset cases).

The results and the discussion are reported separately for each variable in the following sections. The assumption of normality of distribution and heteroscedasticity of the analysed scores were tested. When not complied, 1 Way ANOVA was conducted on main effects, adopting Welch and Brown - Forsythe tests. The resulting probabilities are reported in the same tables in the last column.

## 3. Results and Discussion

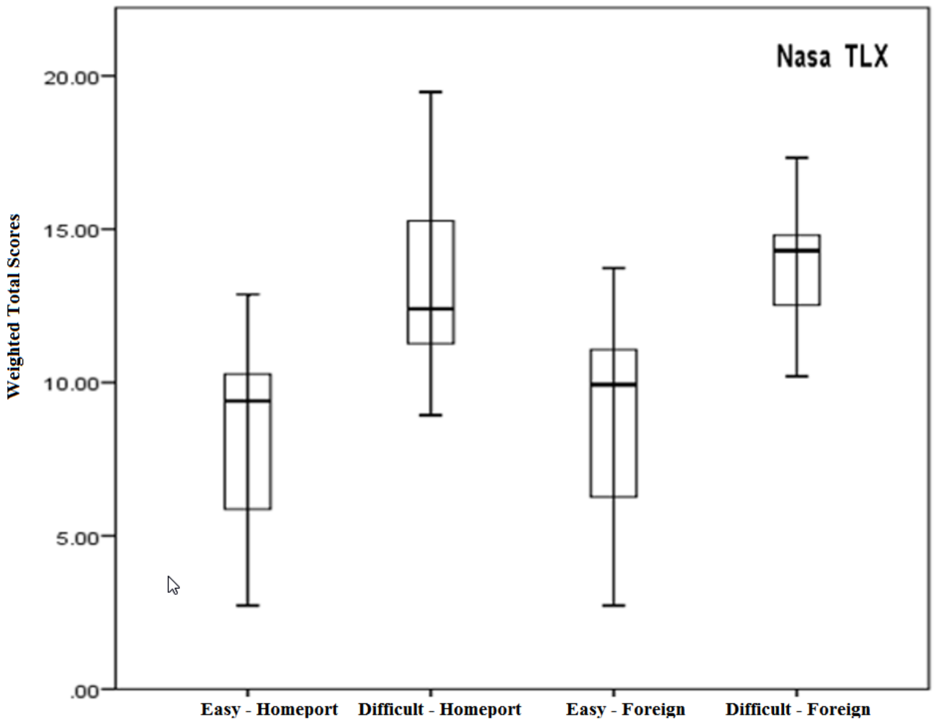
### *3.1 Missing Data*

During the simulations not all the runs were completed by the pilots. A total of four crashes were recorded (three during the berthings in the virtual port and one in the homeport). A "crash" was an impact or a grounding that required the interruption of the simulation and data collection. All the crashes were experienced during the berthings at level 1 of difficulty. Three impacts were also experienced (one in the virtual port with difficulty level 0 and two during the swing in both ports at difficulty level 1). An "impact" was classified as a contact of the vessel with another ship or port infrastructures that did not impede the continuation of the berthing.

### *3.2 NASA TLX*

In figure 2 the boxplots reported, describe the distribution of the NASA TLX scores recorded in the different berthings.

**Figure 2. NASA TLX scores obtained in the four experimental berthings.**




The NASA TLX was adopted as a whole measurements of the engagement or workload experienced by the pilots with reference to the berthings just completed. The factor "phase", hence, could not be examined, and therefore only the factors "port" and "difficulty" were tested with NASA TLX. In addition, the analysis of the 6 different sub scales, identified that the significant results were obtained in the subscales *mental demand*, *performance*, *effort* and *frustration* for the variable "difficulty". *Temporal demand* and *physical demand* were non-significant for the variable "difficulty", as detailed in table 3.

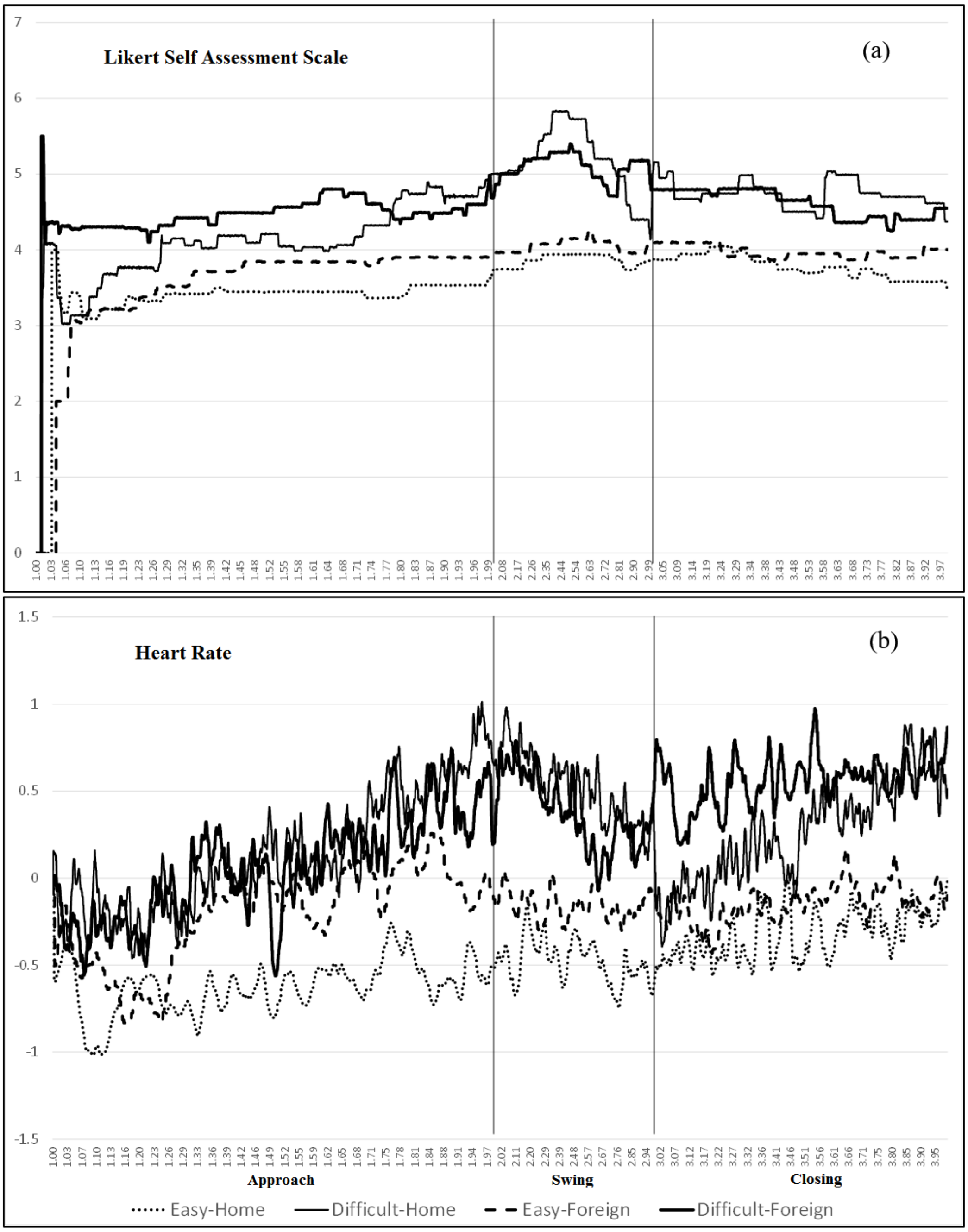
**Table 3. ANOVA results summary table – NASA TLX.**

| Variable | Subscale or Parameter | Significant Factor | 2 Way ANOVA Results | Compared means | Robust Test Results |
|---|---|---|---|---|---|
| **NASA TLX** | Mental Demand | Difficulty | $F(1, 36) = 13.24$, $p = 0.001$ | Easy = 40.35 vs Difficult = 60.25 | $p = 0.001$ (3) $p = 0.001$ (4) |
| | Performance | Difficulty | $F(1, 36) = 5.22$, $p = 0.03$ | Easy = 36.15 vs Difficult = 56.60 | $p = 0.03$ (3) $p = 0.03$ (4) |
| | Effort | Difficulty | $F(1, 36) = 3.89$, $p = 0.056$ (1) | Easy = 25.65 vs Difficult = 38.25 | $p = 0.051$ (3) $p = 0.051$ (4) |
| | Frustration | Difficulty | $F(1, 36) = 4.49$, $p = 0.04$ | Easy = 10.45 vs Difficult = 19.65 | $p = 0.04$ (3) $p = 0.04$ (4) |
| | *Temporal Demand* | *None* (5) | *$F(1, 36) = 0.18$, $p = 0.67$* | *Easy = 19.80 vs Difficult = 22.65* | *$p = 0.66$ (3) $p = 0.66$ (4)* |
| | *Physical Demand* | *None* (5) | *$F(1, 36) = 0.21$, $p = 0.65$* | *Easy = 2.15 vs Difficult = 3.35* | *$p = 0.64$ (3) $p = 0.64$ (4)* |
| | **Total Rating** | Difficulty | $F(1, 38) = 21.12$, $p = 0.001$ | Easy = 8.94 vs Difficult = 13.48 | $p = 0.001$ (3) $p = 0.001$ (4) |
| **Notes** | (1) Only marginal; (2) Tuckey HSD Test; (3) Welch; (4) Brown – Forsythe; (5) *Non significant results provided for the factor Difficulty* | | | | |

The workload was mainly felt as mental and related to the achievement of a desired result, i.e. matching as much as possible the initially provided passage plan. Pilots were also less satisfied with their level of performance in the more difficult berthings. They exerted more effort in the attempt to maintain good control of the vessel during the more difficult scenarios, experiencing a higher level of frustration. No significant interaction between the factors "difficulty" and "port" was found.

### *3.3 Self Assessment Likert Scale*

**Figure 3. Time comparison between Likert scale scores (a) and HR (b) obtained in the four berthings. Vertical lines separate the different phases.**




In Figure 3 it is possible to directly compare the results obtained in the self assessment Likert scale (figure 3(a)) with those obtained in the heart rate (figure 3(b)). For each subject, the scores collected during each berthing were divided into the three phases of the execution. The timestamp associated to each recorded score, was transformed in a fraction of unit, considering the duration of each single phase as a whole unit of time (Approach - from 1 to 2, Swing - from 2 to 3, Closing - from 3 to 4). The graphs were obtained running a moving average through the results recorded from all the ten subjects, in each berthing. The results obtained using the self-assessment Likert scale further supported the NASA TLX results. It can be noticed how, consistently with what already reported in the NASA TLX results, The difficult berthings (continuous lines in 3(a) and 3(b)) reported higher values compared to the easy berthings (dotted lines in 3(a) and 3(b)). Through a more direct and continuous measurement on seven levels of pilots' engagement throughout the berthings, a significant difference between easy and difficult berthings was confirmed. Moreover, it was also possible to highlight a slightly increased evaluation of the workload experienced during the berthings performed in the virtual port, compared to those performed in the homeport, as detailed in table 4.

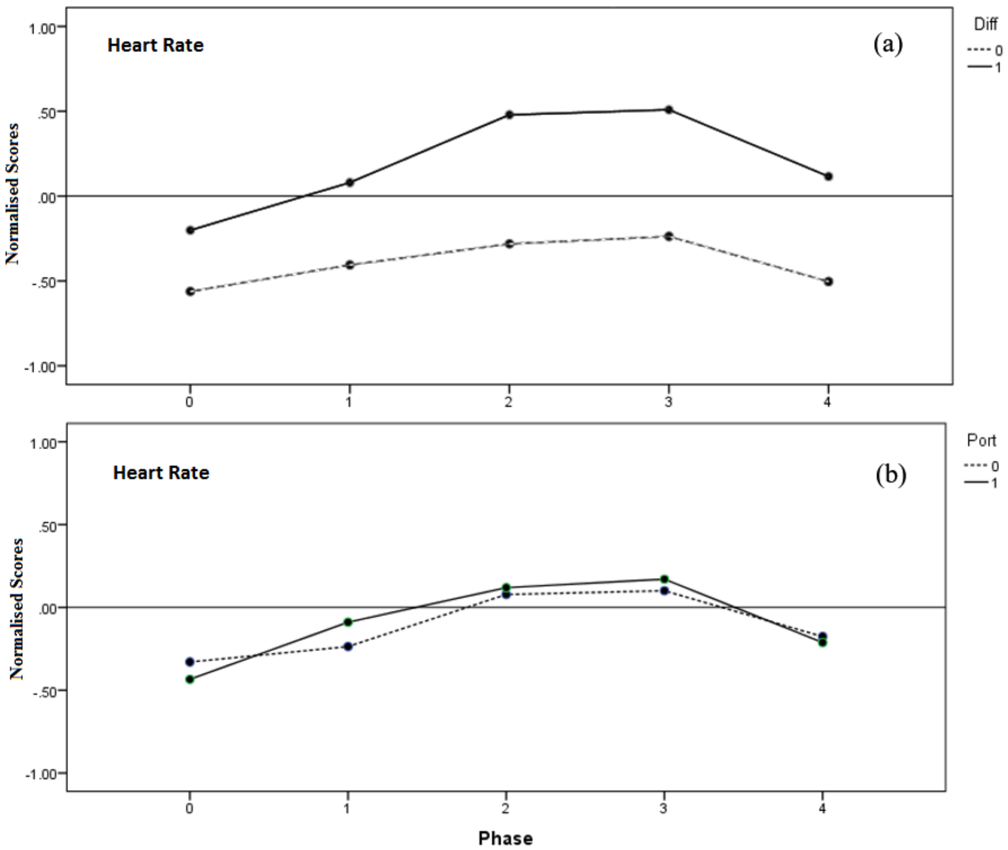
**Table 4. ANOVA results summary table – Self Assessment Likert Scale.**

| Variable | Significant Factor | 3 Way ANOVA Results | Compared means | Robust Test Results |
|---|---|---|---|---|
| **Self Assessment Scale** | Difficulty | $F(1, 102) = 44.06$, $p = 0.001$ | Easy = 3.84 vs Difficult = 4.68 | $p = 0.001$ (3) $p = 0.001$ (4) |
| | Port | $F(1, 102) = 4.95$, $p = 0.03$ | Home = 4.12 vs Foreign = 4.40 | $p = 0.07$ (3) $p = 0.07$ (4) |
| | Phase | $F(2, 102) = 7.00$, $p = 0.001$ | Approach = 3.99; Swing = 4.56; Closing = 4.26; | $p = 0.02$ (3) $p = 0.01$ (4) |
| | Phase (Post Hoc) | $p = 0.001$ (2) | Approach = 3.99 vs Swing = 4.56 | |
| **Notes** | (1) Only marginal; (2) Tuckey HSD Test; (3) Welch; (4) Brown - Forsythe | | | |

According to pilots, the workload experienced during the different "phases" was different, justifying the adoption of such additional factor to study the effect of the experimental conditions on the physiological variables collected. The significant difference was specifically experienced between the phases "approach" and "swing". No significant interactions were recorded between the factors in the 3 Way ANOVA, so only the main effects are reported in table 4.

### *3.4 ECG – HR and LF/HF*

**Figure 4. ECG - Heart Rate (HR). Comparison of the normalized means along the factor "phase" for different levels of "difficulty" (a) and in different "ports" (b).**




Results obtained in the 3 Way ANOVA of the heart rate, were consistent with the results obtained in the Likert scale and in the NASA TLX. Measures of the HR showed a clear increment in the recordings during difficult berthings (3 Way main effect on factor "difficulty"), as depicted in the gap reported in figure 4(a). Graphs provided in figure 4(a) and 4(b) highlight also how the HR showed a consistent pattern throughout the succession of the 5 phases. Obtained means started with lower values during the phase "baseline pre" (number 0 in abscissa, with ordinate values around - 0.5). HR increased during the "approach" (number 1 in abscissa) and even more during the phases "swing" (number 2 in abscissa) and "closing" (number 3 in abscissa). A post hoc analysis revealed that the differences between "baseline pre" and "swing" and between "baseline pre" and "closing" were significant (3 Way main effect on factor "phase"). The final phase (number 4), when a "baseline post" was collected, showed a decrement in scores even if not statistically significant compared to the other phases. In figure 4(a) it is also possible to notice how the means obtained in the two baseline phases (0 and 4) are different between the two levels of difficulty. We think that such difference could have been originated from some anticipatory stress for phase 0. The pilots were told before the recording of the "baseline pre" what was the berthing that was just about to begin. Pilots were not made aware about the whole berthing order for the day, until they actually executed the exercises. The increased value in phase 4 in the difficult berthings could probably be caused by: (1) a general increases response (2) the crashes experienced, that would have had phase 4 recorded in the five minutes immediately following such event. The significant results are reported in table 5.

**Table 5. ANOVA results summary table – ECG.**

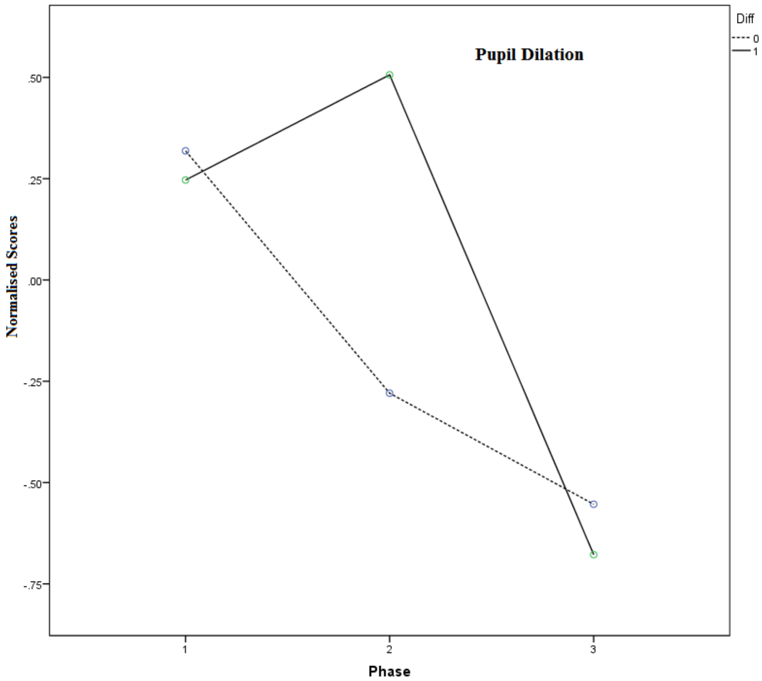
| Variable | Subscale Or Parameter | ANOVA Type - Significant Factor | ANOVA Results | Compared means | Diff % | Robust Test Results |
|---|---|---|---|---|---|---|
| ECG | Heart Rate | 3 Way - Difficulty | F(1, 161) = 49.90, p = 0.001 | Easy = -0.40 vs Difficult = 0.20 | + 23 | p = 0.001 (3) p = 0.001 (4) |
| | | 3 Way - Phase | F(4, 161) = 5.35, p = 0.001 | BLine Pre = -0.38; Approach = -0.16; Swing = 0.10; Closing = 0.14; BLine Post = -0.19 | | p = 0.01 (3) p = 0.01 (4) |
| | | 3 Way - Phase (Post Hoc) | p = 0.001 (2) | BL Pre = -0.38 vs Swing = 0.10 | + 19 | |
| | | | | BL Pre = -0.38 vs Closing = 0.14 | + 20 | |
| | | 3 Way - Diff X Port | F(1, 161) = 7.45, p = 0.01 | | | |
| | | 2 Way - Diff X Port (Diff = 0) | F(1, 161) = 4.99, p = 0.027 (1) | Home = -0.52 vs Frgn= -0.27 | + 9 | |
| | | 2 Way - Diff X Port (Port = 0) | F(1, 161) = 49.76, p = 0.001 | Easy = -0.52 vs Diff = 0.30 | + 31 | |
| | | 2 Way - Diff X Port (Port = 1) | F(1, 161) = 9.07, p = 0.003 | Easy = -0.27 vs Diff = 0.09 | + 14 | |
| | LF / HF | Nil | | | | |
| **Notes** | (1) Only marginal; (2) Tuckey HSD Test; (3) Welch; (4) Brown - Forsythe | | | | | |

The 3 Way ANOVA of the HR highlighted also an interaction between factors “difficulty” and “port”. Additional 2 Way ANOVA (with Bonferroni adjustment) were performed using the factors “difficulty” and “port” and collapsing the factor “phase”. A marginally significant simple main effect (p = 0.027) was found considering the easy berthings (Diff = 0), where the HR was higher in the foreign port compared to the homeport. A possible explanation could reside in the fact that of the four crashes recorded during level 1 of difficulty, three took place in the virtual port. It could be argued that the data missing due to the interruption of the recordings, could have raised the average of the scores above those collected in the homeport. The 2 Way ANOVA confirmed higher scores of HR for the difficult berthings, in both ports.

From the analysis of the HRV obtained from the ECG signal, the scores of the ratio LF/HF were also considered, but no significant difference between the means was observed.

### *3.5 Pupil Dilation*

**Figure 5. Normalized pupil dilation. Interaction between factors "difficulty" and "phase".**




For the pupil dilation it was not possible to record a pre and post baseline, since the eye trackers were not active during those resting phases. In table 6 it can be noticed the significant effect on the main factor "phase", that showed higher values during the "approach" and during the "swing" compared to the "closing".

**Table 6. ANOVA results summary table – Pupil Dilation.**

| Variable | ANOVA Type – Significant Factor | ANOVA Results | Compared means | Diff % | Robust Test Results |
|---|---|---|---|---|---|
| Pupil Dilation | 3 Way - Difficulty | F(1, 100) = 3.28, p = 0.073 (1) | Home = - 0.17 vs Foreign = - 0.02 | + 2 | p = 0.07 (3) p = 0.07 (4) |
| | 3 Way - Phase | F(2, 100) = 25.19, p = 0.001 | Approach = 0.28; Swing = 0.11; Closing = -0.62 | | p = 0.001 (3) p = 0.001 (4) |
| | 3 Way - Phase (Post Hoc) | p = 0.001 (2) | Approach = 0.28 vs Closing = -0.62; | - 34 | |
| | | p = 0.001 (2) | Swing = 0.11 vs Closing = -0.62 | - 28 | |
| | 3 Way - Diff X Phase | F(2, 100) = 7.37, p = 0.001 | | | |
| | 2 Way - Diff X Phase (Phase = 2) | F(1, 100) = 17.44, p = 0.001 | Easy = -0.28 vs Diff = 0.51 | + 30 | |
| | 2 Way - Diff X Phase (Diff = 0) | F(2, 100) = 12.26, p = 0.001 | Approach = 0.32 vs Swing = -0.28; | - 24 | |
| | | | Approach = 0.32 vs Closing = -0.55 | - 33 | |
| | 2 Way - Diff X Phase (Diff = 1) | F(2, 100) = 18.80, p = 0.001 | Approach = 0.25 vs Closing = -0.68; | - 35 | |
| | | | Swing = 0.51 vs Closing = -0.68 | - 45 | |
| Notes | (1) Only marginal; (2) Tuckey HSD Test; (3) Welch; (4) Brown - Forsythe | | | | |

This might suggest a higher level of response elicited by expectations related to the exercise just started. The effect that was particularly interesting, though, was the interaction between the factors "phase" and "difficulty". 2 Way ANOVA were conducted to investigate further. These analysis confirmed a significant difference between easy and difficult berthings during the "swing" (Phase = 2). It was also possible to confirm how, in the easy berthings (Diff = 0), the mean obtained in the "approach" was significantly higher than the other two phases. In the difficult berthings (Diff = 1), instead, the "closing" was significantly lower than the other two. Observing figure 5 it can be noticed how there was a drop of values in the easier berthings compared to the most difficult ones. The difficult berthings showed, in phase 2, the absolute highest scores for this physiological variable. This result is also reflected in the marginal main effect obtained on the factor "difficulty".

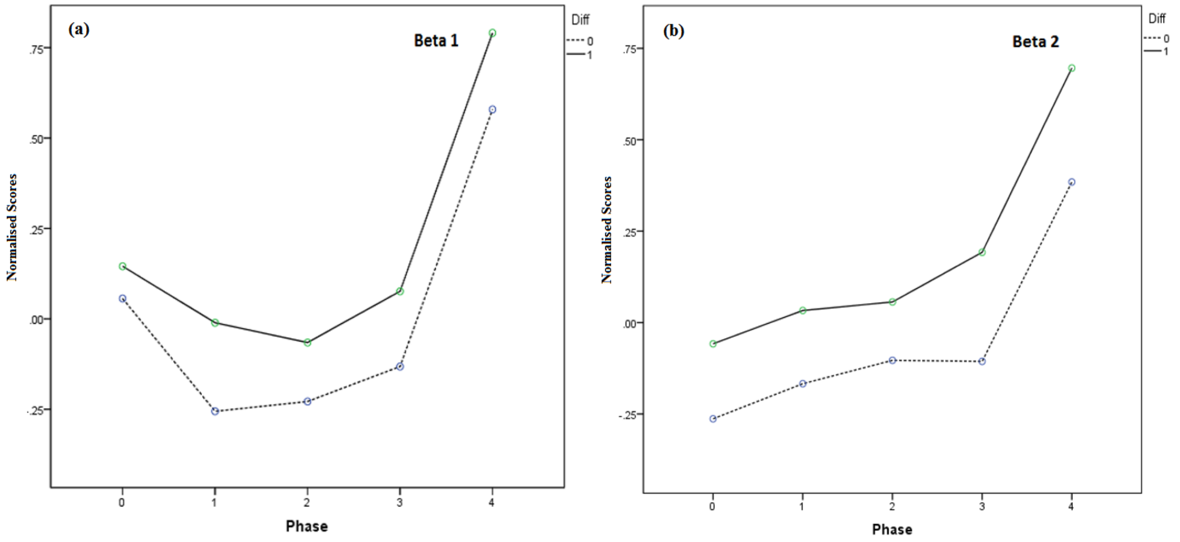
### *3.6 EEG – Bands Beta 1 and Beta 2*

**Table 7. ANOVA results summary table – EEG.**

| Variable | Subscale Or Parameter | Significant Factor | ANOVA Results | Compared means | Diff % | Robust Test Results |
|---|---|---|---|---|---|---|
| EEG | Beta 1 | 3 Way - Difficulty | F(1, 152) = 4.04,<br>p = 0.05 | Easy = 0.00 vs<br>Difficult = 0.19 | + 8 | p = 0.43 (3)<br>p = 0.43 (4) |
| | | 3 Way - Phase | F(4, 152) = 10.66,<br>p = 0.001 | BLine Pre = 0.10;<br>Approach = -0.13;<br>Swing = -0.15;<br>Closing = -0.03;<br>BLine Post = 0.68 | | p = 0.001 (3)<br>p = 0.001 (4) |
| | | 3 Way – Phase (Post Hoc) | p = 0.001 (2) | BL Post = 0.68 > Other Phases | ~+ 21 | |
| | Beta 2 | 3 Way – Difficulty | F(1, 152) = 6.86,<br>p = 0.01 | Easy = -0.05 vs<br>Difficult = 0.18 | + 9 | p = 0.01 (3)<br>p = 0.01 (4) |
| | | 3 Way – Phase | F(4, 152) = 7.11,<br>p = 0.001 | BLine Pre = -0.16;<br>Approach = -0.07;<br>Swing = -0.02;<br>Closing = 0.04;<br>BLine Post = 0.54 | | p = 0.001 (3)<br>p = 0.001 (4) |
| | | 3 Way – Phase (Post Hoc) | p = 0.001 (2) | BL Post = 0.54 > Other Phases | ~+ 19 | |
| | | 3 Way - Diff X Port | F(1, 152) = 4.87,<br>p = 0.03 | | | |
| | | 2 Way - Diff X Port (Port = 0) | F(1, 152) = 11.94,<br>p = 0.001 | Easy = -0.18 vs<br>Diff = 0.26 | + 17 | |
| | | 2 Way - Diff X Port (Diff = 0) | F(1, 152) = 3.92,<br>p = 0.045 (1) | Home = -0.18 vs<br>Frgn= 0.07 | + 10 | |
| **Notes** | (1) Only marginal; (2) Tuckey HSD Test; (3) Welch; (4) Brown - Forsythe | | | | | |

Results obtained in those two bands, as reported in table 7, confirmed as main effects higher physiological response in the more difficult berthings. Such response seemed to be carried over beyond the duration of the berthings, showing the highest scores when the baseline post was recorded compared to all the other phases.

**Figure 6. EEG - Comparison of the normalized means in band Beta 1 (a) and in band Beta 2 (b) along the factor "phase" for different levels of "difficulty".**




In the 3 Way ANOVA, an interaction between the factors "difficulty" and "port" was found in the variable Beta 2. 2 Way ANOVA were conducted. Significant results were found in the homeport (Port = 0) with the easy berthings scoring less than the difficult ones. The other significant result was only relative to the easy berthings (Diff = 0): scores in the foreign port were higher than the homeport. It has to be highlighted though that the adoption of a wireless device, gaining contact with the scalp only through pads not glued but soaked in saline solution, might have not granted an optimal conductivity of the signal during the execution of the berthings. Signals might have been disturbed by face muscle contractions artefacts and the movement of participants during the exercises. Freedom of movement was instead reduced during the collection of the baselines. A more realistic comparison could be then achieved simply comparing the results coming from the "baselines pre" and "post" (phase numbers 0 and 4). In this case it can be noticed how there was a significant increment of the elicited physiological response between before and after the berthings.

### *3.7 Correlations*

Pearson correlation coefficients were calculated between all the variables. Complete results are reported in table 8.

**Table 8. Correlations between dependent variables**

| | | Nasa TLX | Self Assess. | HR | LF/HF | Pupil Dilation | EEG_B1 | EEG_B2 |
|---|---|---|---|---|---|---|---|---|
| Nasa TLX | Pearson Correlation | 1 | .645** | .157 | -.143 | .184 | -.032 | -.002 |
| | Sig. (2-tailed) | | .001 | .103 | .138 | .052 | .751 | .987 |
| | N | 120 | 114 | 109 | 109 | 112 | 104 | 104 |
| Self Asses. | Pearson Correlation | .645** | 1 | .334** | .014 | .243** | .101 | .167 |
| | Sig. (2-tailed) | .001 | | .001 | .882 | .010 | .306 | .090 |
| | N | 114 | 114 | 109 | 109 | 112 | 104 | 104 |
| HR | Pearson Correlation | .157 | .334** | 1 | -.122 | -.011 | .283** | .228* |
| | Sig. (2-tailed) | .103 | .001 | | .205 | .907 | .004 | .022 |
| | N | 109 | 109 | 109 | 109 | 109 | 101 | 101 |
| LF/HF | Pearson Correlation | -.143 | .014 | -.122 | 1 | -.351** | .035 | .017 |
| | Sig. (2-tailed) | .138 | .882 | .205 | | .001 | .656 | .831 |
| | N | 109 | 109 | 109 | 178 | 109 | 165 | 165 |
| Pupil Dilation | Pearson Correlation | .184 | .243** | -.011 | -.351** | 1 | -.200* | -.105 |
| | Sig. (2-tailed) | .052 | .010 | .907 | .001 | | .042 | .288 |
| | N | 112 | 112 | 109 | 109 | 112 | 104 | 104 |
| EEG_B1 | Pearson Correlation | -.032 | .101 | .283** | .035 | -.200* | 1 | .887** |
| | Sig. (2-tailed) | .751 | .306 | .004 | .656 | .042 | | .001 |
| | N | 104 | 104 | 101 | 165 | 104 | 172 | 172 |
| EEG_B2 | Pearson Correlation | -.002 | .167 | .228* | .017 | -.105 | .887** | 1 |
| | Sig. (2-tailed) | .987 | .090 | .022 | .831 | .288 | .001 | |
| | N | 104 | 104 | 101 | 165 | 104 | 172 | 172 |

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

It is possible to observe how the HR was the physiological variable correlating most and significantly (r = .334 with p <= .01 - 2 tailed) with the Likert scale. To gain a better understanding about the relationship between HR and Likert scale and its meaning, we can refer to figure 3. In figure 3(a) it is presented the averaged value across pilots of the experienced mental workload as reported using the seven levels Likert scale. It can be noticed how, in all the berthings, there is a progressive increase of such measurement throughout the "approach". The increase continues until half of the "swing", when the self-reported workload reaches its climax. The swing phase represented the moment in the berthing where the ship had to be rotated 180 degrees from her initial course. This was obtained through the intense use of tugs, controlled by the pilot via radio communications. The control of the vessel in terms of her position and speed had to be extremely accurate since any error would have been very difficult to recover, leading to an impact with the surrounding port features. Half of the swing was the moment when the pilot could clearly appreciate if the induced swing would have been safely concluded or it would have reached an undesired outcome. In figure 3(b) it is possible to notice how the HR basically followed the experienced mental workload even if it started to drop since the beginning of the "swing" phase. Also the pupil dilation recorded a lower but significant correlation ($r = .243$ with $p <= .01$ - 2 tailed). Those results may suggest that the use of those variables may be further investigated, in the research of independent indicators that could provide indirect measurements of mental workload.

## 4. Conclusions

This paper reports results related to mental workload and physiological responses obtained from a group of ten expert pilots while performing berthings in a shipping simulator. Several physiological variables were collected and analysed in order to obtain measurements that could be compared to scores from NASA TLX and a second self-assessment workload Likert scale.

Results obtained from measuring ECG, EEG, and pupil dilation provided some indications that physiological variables correlated to scores obtained from self-assessment scales. Light correlations were highlighted specifically between the self-assessment Likert scale and heart rate ($r = .334$) and pupil dilation ($r = .243$). Increasing the level of difficulty induced a significant increment in the levels of responses, particularly in the HR (23% of increment).

In this regard, the part of the berthing that elicited the strongest reaction was the swing (best depicted in figure 5, which graphs pupil dilation responses). In that figure a significant interaction between the factors "difficulty" and "phase" is depicted. Controlling the safe rotation of a large vessel in constrained waters and in critical environmental conditions, challenged even expert pilots, and this was consistently shown not only in pilot's verbal reports, but also by all their physiological responses.

The inclusion of a novel or unknown port in the research protocol did not show a statistical significant effect of increasing the experienced workload. Marginal increases in self-assessed workload were not reflected in similar changes in the physiological data. This result may initially suggest that the use of pilots own port did not offer such a major advantage in the way berthings were performed. It has to be noticed, though, that four crashes happened during the conduction of the difficult berthings (preventing the complete collection of scores in those berthings) and three of these crashes occurred in the virtual port (affecting the mean of the scores collected during those berthings).

One of the generally accepted "wisdoms" offered for mandating the use of marine pilots is that they have expert "local" knowledge. While the existence of this knowledge is not contested, we suggest that marine pilots also have generalised expertise associated with ship-handling that supports shiphandling performance in any port and these skills may well interact with the "local knowledge" effect we might have discovered by manipulating this variable in our experimental protocol. We view this outcome as useful, because it will guide the future manipulation of the variables used to differentiate the "easy" and "difficult" conditions. This will help move us some of the way towards identifying an upper "redline" (M. S. Young et al., 2015) for the task demands of marine pilots, in the context of available resources, although we acknowledge that this is extremely difficult to identify.

By manipulating spatial constraints, environmental conditions, vessel characteristics, tugs, and traffic interactions we created an ecologically valid condition in which the majority of pilots completed the berthing, but several pilots could not. This was associated with significant increases in workload during the swing, correlated with physiological variables as described above. Our method and results indicate that the expert pilot is more often than not able to perform berthings at the limits of what was possible, where the mental workload included holding at least the following information in working memory:

- spatial constraints (distances from dangers, perception and correction of vessel motion);
- environmental conditions (effects on the ship movement of wind, current, hydrodynamic forces, as they also constantly vary according to vessel motion);
- vessel characteristics (manoeuvrability capabilities and limitations, effective use of vessel propulsion and steerage);
- tugs (tugs call signs, location relative to main ship, current tugs power usage, spare tugs capacity),
- traffic interactions (other means operating in the port and representing a potential danger for the safety of the berthing).

In order to continue the analysis of marine pilotage workload, future research will need to quantify measures of performance and to consider the issue of resource supply and demand. Resource supply and demand for marine pilots is an issue that has attracted debate within the maritime industry for a number of reasons. There has been some argument as to whether the embarkation of a pilot increases a bridge team by one or reduces the team to one. The erosion of the competency of seafarers to support a pilot during the berthing has been questioned (Goodfellow, 2008). Tug masters have also traditionally simply followed verbal orders, yet they are mostly well trained ship handlers in their own right, and have the potential to play a more prominent role. The Vessel Traffic Service (VTS) (a maritime equivalent of Air Traffic Control) has almost unilaterally played a minor role in the process of moving ships, but has the potential to support ship movements through the management of a much broader ‘traffic picture’. Each of these people (ships’ bridge teams, tug masters, Vessel Traffic Service operators) might provide significant opportunities in the external resource supply side to reduce the task demands imposed on marine pilots, especially during berthings in critical conditions.

In summary, the results showed how a full mission bridge simulator can be profitably exploited to create different scenarios inducing different levels of engagement or mental workload, and those different levels of workload are also correlated to physiological responses. Those responses can be used as unobtrusive, indirect measurements of mental workload, allowing pilots to concentrate on their work performance and not on data collection.

### *4.1 Future Applications and Added Value.*

This research creates an experimental platform that will be able to identify when changes in physiological responses indicate higher or critical levels of mental workload for marine pilots. Such a method may be useful in operational contexts where change occurs such as when a new ship type is to commence operations at the port or when the port decides to develop new berths or adjust approaches/channel dimensions. The method may also useful to designers of equipment to assess the effect of their equipment on tasks, and therefore workload.

This research has moved us one step closer to understanding a level of workload or threshold that, once exceeded, could expose pilots’ performance to detrimental results, or in other words, where the “red line” would be crossed (Brookhuis et al., 2003; M. S. Young et al., 2015). However, this issue is complex, related to multiple variables, including individual variation and is highly context dependent. Redlines become even more difficult to determine if they are considered from the perspective of physiological variables. More research is needed around this issue, and the problem may well turn out to be intractable.

In order to better understand and define what could be considered an “optimal range of activity” additional variables should be collected. The experimental design in this study does not account for information associated with auditory or visual perception or verbal linguistic activity. However we used eye trackers to monitor the pupil dilation. This device was collecting gaze information as well as voice recordings.

Through the gaze analysis we would be able to obtain information about gaze distribution and fixation, (and therefore attention) specifically with reference to the source of information (electronic equipment, external visual aids..) preferred by pilots (Itoh et al., 1990). This could offer an important insight about information resource management and shedding preferences once task demand begins to overcome pilot capabilities (M. S. Young et al., 2015). Voice recordings would provide data such as the number and type of communications or orders per minute given by the pilots (Luca Orlandi et al., 2014). To execute their plan, pilots give orders to ships' crew and tug boats. We could hypothesise that during more complex and challenging scenarios, a substantial increase in the number of orders or communications per minute and / or a change in gaze behaviour would be recorded, relating to different results in physiological variables and performance outcomes (Luca Orlandi et al., 2015).

The current research was conducted only in a simulated environment. Future developments should consider the collection and comparison of similar data in a real working environment. This does, however, pose significant challenges. For real-world data collection, it will be critical for instruments to be as resilient and unobtrusive as possible, in order not to distract or interfere with berthing operations and also to collect data without failures, allowing stronger statistical analysis and results. Should that data collection be possible, it could provide a better understanding of normal and abnormal, personal and group response levels, which could help to identify critical operations and levels of performance. Once those levels were identified, they could be exploited as prodromal indicators of the developing of critical conditions. Beyond this, such data collection will provide us with an understanding of the realism of the simulated environment through a comparative analysis of the workload and physiological data.

Finally, we anticipated that this study could inform improvements in Australian national pilotage standards (ATC, 2008) around issues such as use of simulation facilities for training, continuing professional development of pilots and the influence of workload on fatigue, as well as broader debates in this industry. Debate continues in the maritime domain regarding issues such as increasing automation within pilotage, use of technology, and even remote pilotage (Brooks et al., 2016), and our method and results could make a contribution associated with these issues by identifying where and how workload changes across a manoeuvre. Improvements in pilotage standards could include using our experimental protocol to update training standards, creating materials for Continuing Professional Development (CPD) of marine pilots around the relationship between workload and stress, and using these results as the basis for further studies which could lead to modifications of fatigue risk management practices when extreme workloads have occurred.

## Acknowledgements

We would like to thank the Australasian Marine Pilots Institute and Smartship Simulator for the priceless support offered in the realization of this work. We are deeply grateful to the Pilots that patiently bore with the intervention of too many obnoxious devices! We would also like to thank Prof Joanne Wood and Dr. Alexander Black (Queensland University of Technology - Faculty of Health, School - Optometry and Vision Science), for providing us with some of those scary tools.

# Appendix A

***Self assessment Likert scale levels.***

| Level | State | Description |
|---|---|---|
| 7 | Extremely Demanding | An extremely demanding situation that it is just about to be out of hand |
| 6 | Very Demanding | A challenging situation that requires the complete attention of the shiphandler, working at almost 100% of his capabilities |
| 5 | Demanding | A situation requiring more attention than normal, but not felt as critical as level 6 |
| 4 | Average | A situation with normal level of involvement where the shiphandler can feel perfectly capable to achieve the desired outcome with a necessary but comfortable level of effort (routine operation) |
| 3 | Easy | An easy situation offering no specific challenge, with required effort below the average |
| 2 | Very Easy | A very comfortable, almost effortless situation |
| 1 | Extremely Easy | A situation of “boredom”, with very little or no involvement at all |

# Appendix B

## *Anova Results*

Dependent Variable: Nasa TLX – Total Weighed Score

| Source | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 632.762[a] | 3 | 210.921 | 25.639 | .000 |
| Intercept | 15079.692 | 1 | 15079.692 | 1833.045 | .000 |
| Diff | 618.348 | 1 | 618.348 | 75.165 | .000 |
| Port | 14.366 | 1 | 14.366 | 1.746 | .189 |
| Diff * Port | .048 | 1 | .048 | .006 | .939 |
| Error | 954.283 | 116 | 8.227 | | |
| Total | 16666.737 | 120 | | | |
| Corrected Total | 1587.045 | 119 | | | |

a. R Squared = .399 (Adjusted R Squared = .383)

Nasa TLX - Subscales

| Source | Dependent Variable | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Corrected Model | Mental Demand | 4097.000[a] | 3 | 1365.667 | 4.565 | .008 |
| | Physical Demand | 40.100[b] | 3 | 13.367 | .200 | .896 |
| | Temporal Demand | 161.675[c] | 3 | 53.892 | .121 | .947 |
| | Performance | 5329.075[d] | 3 | 1776.358 | 2.214 | .103 |
| | Effort | 1388.100[e] | 3 | 462.700 | 1.336 | .278 |
| | Frustration | 925.700[f] | 3 | 308.567 | 1.638 | .198 |
| Intercept | Mental Demand | 101203.600 | 1 | 101203.600 | 338.304 | .000 |
| | Physical Demand | 302.500 | 1 | 302.500 | 4.516 | .041 |
| | Temporal Demand | 18020.025 | 1 | 18020.025 | 40.335 | .000 |
| | Performance | 86025.625 | 1 | 86025.625 | 107.226 | .000 |
| | Effort | 42120.100 | 1 | 42120.100 | 121.639 | .000 |
| | Frustration | 9060.100 | 1 | 9060.100 | 48.091 | .000 |
| Diff | Mental Demand | 3960.100 | 1 | 3960.100 | 13.238 | .001 |
| | Physical Demand | 14.400 | 1 | 14.400 | .215 | .646 |
| | Temporal Demand | 81.225 | 1 | 81.225 | .182 | .672 |
| | Performance | 4182.025 | 1 | 4182.025 | 5.213 | .028 |
| | Effort | 1345.600 | 1 | 1345.600 | 3.886 | .056 |
| | Frustration | 846.400 | 1 | 846.400 | 4.493 | .041 |
| Port | Mental Demand | 122.500 | 1 | 122.500 | .409 | .526 |
| | Physical Demand | .100 | 1 | .100 | .001 | .969 |
| | Temporal Demand | 46.225 | 1 | 46.225 | .103 | .750 |
| | Performance | 731.025 | 1 | 731.025 | .911 | .346 |

| Source | Dependent Variable | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| | Effort | 2.500 | 1 | 2.500 | .007 | .933 |
| | Frustration | 72.900 | 1 | 72.900 | .387 | .538 |
| Diff * Port | Mental Demand | 14.400 | 1 | 14.400 | .048 | .828 |
| | Physical Demand | 25.600 | 1 | 25.600 | .382 | .540 |
| | Temporal Demand | 34.225 | 1 | 34.225 | .077 | .784 |
| | Performance | 416.025 | 1 | 416.025 | .519 | .476 |
| | Effort | 40.000 | 1 | 40.000 | .116 | .736 |
| | Frustration | 6.400 | 1 | 6.400 | .034 | .855 |
| Error | Mental Demand | 10769.400 | 36 | 299.150 | | |
| | Physical Demand | 2411.400 | 36 | 66.983 | | |
| | Temporal Demand | 16083.300 | 36 | 446.758 | | |
| | Performance | 28882.300 | 36 | 802.286 | | |
| | Effort | 12465.800 | 36 | 346.272 | | |
| | Frustration | 6782.200 | 36 | 188.394 | | |
| Total | Mental Demand | 116070.000 | 40 | | | |
| | Physical Demand | 2754.000 | 40 | | | |
| | Temporal Demand | 34265.000 | 40 | | | |
| | Performance | 120237.000 | 40 | | | |
| | Effort | 55974.000 | 40 | | | |
| | Frustration | 16768.000 | 40 | | | |
| Corrected Total | Mental Demand | 14866.400 | 39 | | | |
| | Physical Demand | 2451.500 | 39 | | | |
| | Temporal Demand | 16244.975 | 39 | | | |
| | Performance | 34211.375 | 39 | | | |
| | Effort | 13853.900 | 39 | | | |
| | Frustration | 7707.900 | 39 | | | |

a. R Squared = .276 (Adjusted R Squared = .215)

b. R Squared = .016 (Adjusted R Squared = -.066)

c. R Squared = .010 (Adjusted R Squared = -.073)

d. R Squared = .156 (Adjusted R Squared = .085)

e. R Squared = .100 (Adjusted R Squared = .025)

f. R Squared = .120 (Adjusted R Squared = .047)

Dependent Variable: Self Assessment Likert Scale

| Source | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 29.672[a] | 11 | 2.697 | 5.980 | .000 |
| Intercept | 2052.522 | 1 | 2052.522 | 4550.308 | .000 |
| Diff | 19.874 | 1 | 19.874 | 44.058 | .000 |
| Port | 2.233 | 1 | 2.233 | 4.951 | .028 |
| Phase | 6.319 | 2 | 3.159 | 7.004 | .001 |
| Diff * Port | .030 | 1 | .030 | .067 | .796 |
| Diff * Phase | 1.188 | 2 | .594 | 1.317 | .273 |
| Port * Phase | .438 | 2 | .219 | .486 | .617 |
| Diff * Port * Phase | .072 | 2 | .036 | .080 | .924 |
| Error | 46.009 | 102 | .451 | | |
| Total | 2122.055 | 114 | | | |
| Corrected Total | 75.682 | 113 | | | |

a. R Squared = .392 (Adjusted R Squared = .327)

Estimated Marginal Means - Self Assessment Likert Scale

| Diff | Port | Phase | Mean | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 0 | 0 | 1 | 3.431 | .212 | 3.010 | 3.853 |
| | | 2 | 3.882 | .212 | 3.461 | 4.303 |
| | | 3 | 3.751 | .212 | 3.330 | 4.173 |
| | 1 | 1 | 3.898 | .212 | 3.476 | 4.319 |
| | | 2 | 4.110 | .212 | 3.689 | 4.531 |
| | | 3 | 3.999 | .212 | 3.578 | 4.420 |
| 1 | 0 | 1 | 4.094 | .212 | 3.673 | 4.515 |
| | | 2 | 4.978 | .224 | 4.534 | 5.422 |
| | | 3 | 4.609 | .224 | 4.165 | 5.053 |
| | 1 | 1 | 4.520 | .212 | 4.099 | 4.941 |
| | | 2 | 5.252 | .224 | 4.808 | 5.696 |
| | | 3 | 4.655 | .254 | 4.152 | 5.159 |

Dependent Variable: ECG - HR

| Source | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 25.215[a] | 19 | 1.327 | 4.223 | .000 |
| Intercept | 1.805 | 1 | 1.805 | 5.743 | .018 |
| Diff | 15.684 | 1 | 15.684 | 49.902 | .000 |
| Port | .024 | 1 | .024 | .077 | .781 |
| Phase | 6.726 | 4 | 1.681 | 5.350 | .000 |
| Diff * Port | 2.341 | 1 | 2.341 | 7.449 | .007 |
| Diff * Phase | 1.081 | 4 | .270 | .860 | .489 |
| Port * Phase | .358 | 4 | .090 | .285 | .887 |
| Diff * Port * Phase | .319 | 4 | .080 | .254 | .907 |
| Error | 50.601 | 161 | .314 | | |
| Total | 79.031 | 181 | | | |
| Corrected Total | 75.817 | 180 | | | |

a. R Squared = .333 (Adjusted R Squared = .254)

Estimated Marginal Means – HR

| Diff | Port | Phase | Mean | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 0 | 0 | 0 | -.606 | .177 | -.956 | -.256 |
| | | 1 | -.602 | .177 | -.952 | -.251 |
| | | 2 | -.398 | .177 | -.748 | -.048 |
| | | 3 | -.337 | .177 | -.687 | .013 |
| | | 4 | -.681 | .187 | -1.050 | -.312 |
| | 1 | 0 | -.518 | .177 | -.869 | -.168 |
| | | 1 | -.210 | .177 | -.560 | .140 |
| | | 2 | -.165 | .177 | -.515 | .185 |
| | | 3 | -.138 | .177 | -.488 | .212 |
| | | 4 | -.327 | .187 | -.696 | .042 |
| 1 | 0 | 0 | -.053 | .177 | -.403 | .298 |
| | | 1 | .129 | .177 | -.221 | .479 |
| | | 2 | .555 | .198 | .163 | .946 |
| | | 3 | .539 | .198 | .147 | .930 |
| | | 4 | .329 | .198 | -.063 | .720 |
| | 1 | 0 | -.350 | .187 | -.719 | .019 |
| | | 1 | .031 | .187 | -.338 | .400 |
| | | 2 | .404 | .198 | .012 | .795 |
| | | 3 | .479 | .229 | .027 | .931 |
| | | 4 | -.097 | .212 | -.516 | .321 |

Dependent Variable: ECG - LF/HF

| Source | Type IV Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 3.905[a] | 19 | .206 | .516 | .953 |
| Intercept | .000 | 1 | .000 | .000 | .985 |
| Diff | .111 | 1 | .111 | .279 | .598 |
| Port | .005 | 1 | .005 | .013 | .908 |
| Phase | 1.558 | 4 | .390 | .978 | .421 |
| Diff * Port | .593 | 1 | .593 | 1.490 | .224 |
| Diff * Phase | .438 | 4 | .109 | .275 | .894 |
| Port * Phase | .221 | 4 | .055 | .139 | .968 |
| Diff * Port * Phase | .964 | 4 | .241 | .605 | .659 |
| Error | 62.929 | 158 | .398 | | |
| Total | 66.836 | 178 | | | |
| Corrected Total | 66.835 | 177 | | | |

a. R Squared = .058 (Adjusted R Squared = -.055)

Estimated Marginal Means - LF/HF

| Diff | Port | Phase | Mean | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 0 | 0 | 0 | .141 | .210 | -.274 | .556 |
| | | 1 | .014 | .200 | -.380 | .408 |
| | | 2 | -.120 | .200 | -.514 | .274 |
| | | 3 | .197 | .200 | -.198 | .591 |
| | | 4 | .219 | .210 | -.197 | .634 |
| | 1 | 0 | .133 | .200 | -.261 | .527 |
| | | 1 | .025 | .200 | -.369 | .419 |
| | | 2 | -.307 | .200 | -.701 | .087 |
| | | 3 | -.084 | .200 | -.478 | .310 |
| | | 4 | .044 | .223 | -.397 | .485 |
| 1 | 0 | 0 | .049 | .200 | -.345 | .444 |
| | | 1 | -.077 | .200 | -.471 | .317 |
| | | 2 | -.212 | .223 | -.653 | .229 |
| | | 3 | .001 | .223 | -.440 | .442 |
| | | 4 | -.147 | .239 | -.619 | .324 |
| | 1 | 0 | -.102 | .210 | -.518 | .313 |
| | | 1 | -.158 | .210 | -.574 | .257 |
| | | 2 | -.009 | .223 | -.449 | .432 |
| | | 3 | .138 | .258 | -.371 | .647 |
| | | 4 | .274 | .239 | -.197 | .745 |

Dependent Variable: Pupil Dilation

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 24.447[a] | 11 | 2.222 | 6.848 | .000 |
| Intercept | .592 | 1 | .592 | 1.825 | .180 |
| Diff | 1.064 | 1 | 1.064 | 3.279 | .073 |
| Port | .572 | 1 | .572 | 1.762 | .187 |
| Phase | 16.352 | 2 | 8.176 | 25.195 | .000 |
| Diff * Port | .174 | 1 | .174 | .537 | .466 |
| Diff * Phase | 4.785 | 2 | 2.393 | 7.373 | .001 |
| Port * Phase | .894 | 2 | .447 | 1.378 | .257 |
| Diff * Port * Phase | 1.110 | 2 | .555 | 1.711 | .186 |
| Error | 32.452 | 100 | .325 | | |
| Total | 57.276 | 112 | | | |
| Corrected Total | 56.899 | 111 | | | |

a. R Squared = .430 (Adjusted R Squared = .367)

Estimated Marginal Means – Pupil DIlation

| Diff | Port | Phase | Mean | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 0 | 0 | 1 | .102 | .180 | -.255 | .460 |
| | | 2 | -.301 | .180 | -.658 | .057 |
| | | 3 | -.651 | .180 | -1.008 | -.293 |
| | 1 | 1 | .535 | .180 | .178 | .893 |
| | | 2 | -.258 | .180 | -.615 | .099 |
| | | 3 | -.456 | .180 | -.814 | -.099 |
| 1 | 0 | 1 | .153 | .180 | -.204 | .510 |
| | | 2 | .296 | .201 | -.104 | .696 |
| | | 3 | -.471 | .201 | -.871 | -.071 |
| | 1 | 1 | .340 | .180 | -.017 | .698 |
| | | 2 | .716 | .190 | .340 | 1.093 |
| | | 3 | -.885 | .215 | -1.312 | -.458 |

Dependent Variable: EEG - Beta 1

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 18.973[a] | 19 | .999 | 2.815 | .000 |
| Intercept | 1.559 | 1 | 1.559 | 4.395 | .038 |
| Diff | 1.431 | 1 | 1.431 | 4.036 | .046 |
| Port | .203 | 1 | .203 | .572 | .451 |
| Phase | 15.123 | 4 | 3.781 | 10.660 | .000 |
| Diff * Port | 1.095 | 1 | 1.095 | 3.087 | .081 |
| Diff * Phase | .132 | 4 | .033 | .093 | .985 |
| Port * Phase | .952 | 4 | .238 | .671 | .613 |
| Diff * Port * Phase | .167 | 4 | .042 | .118 | .976 |
| Error | 53.910 | 152 | .355 | | |
| Total | 73.994 | 172 | | | |
| Corrected Total | 72.883 | 171 | | | |

a. R Squared = .260 (Adjusted R Squared = .168)

Estimated Marginal Means – Beta 1

| Diff | Port | Phase | Mean | Std. Error | 95% Confidence Interval Lower Bound | 95% Confidence Interval Upper Bound |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | .029 | .199 | -.363 | .421 |
| | | 1 | -.336 | .199 | -.728 | .056 |
| | | 2 | -.294 | .199 | -.686 | .098 |
| | | 3 | -.237 | .199 | -.630 | .155 |
| | | 4 | .285 | .211 | -.131 | .701 |
| | 1 | 0 | .084 | .199 | -.309 | .476 |
| | | 1 | -.175 | .199 | -.567 | .217 |
| | | 2 | -.163 | .199 | -.555 | .230 |
| | | 3 | -.026 | .199 | -.418 | .366 |
| | | 4 | .873 | .225 | .428 | 1.318 |
| 1 | 0 | 0 | .179 | .188 | -.193 | .551 |
| | | 1 | .125 | .188 | -.247 | .497 |
| | | 2 | .043 | .211 | -.373 | .459 |
| | | 3 | .093 | .211 | -.323 | .509 |
| | | 4 | .725 | .211 | .309 | 1.141 |
| | 1 | 0 | .112 | .199 | -.281 | .504 |
| | | 1 | -.146 | .199 | -.538 | .246 |
| | | 2 | -.174 | .211 | -.590 | .242 |
| | | 3 | .059 | .225 | -.386 | .504 |
| | | 4 | .857 | .211 | .441 | 1.273 |

Dependent Variable: EEG - Beta 2

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 15.153[a] | 19 | .798 | 2.324 | .003 |
| Intercept | .751 | 1 | .751 | 2.189 | .141 |
| Diff | 2.353 | 1 | 2.353 | 6.857 | .010 |
| Port | .113 | 1 | .113 | .330 | .566 |
| Phase | 9.755 | 4 | 2.439 | 7.108 | .000 |
| Diff * Port | 1.672 | 1 | 1.672 | 4.872 | .029 |
| Diff * Phase | .145 | 4 | .036 | .106 | .980 |
| Port * Phase | .676 | 4 | .169 | .492 | .741 |
| Diff * Port * Phase | .625 | 4 | .156 | .456 | .768 |
| Error | 52.152 | 152 | .343 | | |
| Total | 67.741 | 172 | | | |
| Corrected Total | 67.305 | 171 | | | |

a. R Squared = .225 (Adjusted R Squared = .128)

Estimated Marginal Means – Beta 2

| Diff | Port | Phase | Mean | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 0 | 0 | 0 | -.250 | .195 | -.636 | .135 |
| | | 1 | -.268 | .195 | -.653 | .118 |
| | | 2 | -.176 | .195 | -.561 | .210 |
| | | 3 | -.268 | .195 | -.654 | .118 |
| | | 4 | .082 | .207 | -.327 | .491 |
| | 1 | 0 | -.276 | .195 | -.662 | .110 |
| | | 1 | -.067 | .195 | -.452 | .319 |
| | | 2 | -.031 | .195 | -.417 | .355 |
| | | 3 | .056 | .195 | -.330 | .441 |
| | | 4 | .686 | .221 | .249 | 1.124 |
| 1 | 0 | 0 | -.067 | .185 | -.433 | .299 |
| | | 1 | .177 | .185 | -.189 | .543 |
| | | 2 | .156 | .207 | -.253 | .565 |
| | | 3 | .329 | .207 | -.080 | .738 |
| | | 4 | .690 | .207 | .281 | 1.099 |
| | 1 | 0 | -.050 | .195 | -.435 | .336 |
| | | 1 | -.111 | .195 | -.496 | .275 |
| | | 2 | -.043 | .207 | -.453 | .366 |
| | | 3 | .055 | .221 | -.382 | .493 |
| | | 4 | .702 | .207 | .293 | 1.111 |

## 6.4. Paper IV

# Interpreting changes in marine pilots' perceptual cycle through gaze detection patterns.

Luca Orlandi [1], Benjamin Brooks [1], Joanne Wood [2] and Alexander Black [2]

(1) *University of Tasmania, Australian Maritime College, National Centre for Ports and Shipping, Launceston, Tasmania 7250, Australia.*
(2) *Queensland University of Technology, Faculty of Health, School of Optometry and Vision Science.*

(E-mail: luca.orlandi@utas.edu.au; Tel: +61 3 6324 3999)

**Abstract**

Marine Pilots conduct vessels in and out of ports around the world every day, relying on their knowledge, experience and their perception of the surrounding environment. This study explored the role of perceptions in shiphandling by analysing visual gaze behaviour of 10 marine pilots engaged in mooring manoeuvres. Four mooring manoeuvres, varied by two levels of difficulty, were completed in a bridge simulator. Eye trackers recorded gaze position. Data was analysed to identify specific gaze behaviours combined in meaningful sequences. These sequences were scored within each manoeuvre. The adopted sequences differed significantly during the manoeuvres, as well as across the two levels of difficulty, demonstrating that pilots had to adopt different screening techniques. Significant changes were found in checks of vessel's position, direction and speed. The development of this methodology will enable further exploration of pilots' expertise and demonstrates how empirical measures can be adopted to quantify shiphandling capabilities.

**Practitioner Summary**

Gaze was investigated in 10 marine pilots during mooring simulations. Gaze behaviour showed significant differences during the execution of different manoeuvres. This method can provide empirical and unobtrusive measurements, which can quantify situation awareness and expertise in this occupational group. Implications for training and operations are identified.

**Keywords**

1. Shiphandling performance 2. Eye trackers 3. Ship Simulator 4. Marine Pilotage 5. Behavioural Markers.

## Introduction

Shipping is the predominant transport mode for the movement of freight in the world, carrying around 90% of the world trade. The maritime industry generates an annual income of over half a trillion US dollars in freight rates, with a worldwide population of 1,187,000 seafarers serving globally on internationally trading merchant ships (ICS, 2016). However, seafarers and ships generally remain effectively invisible to both the general public and the research community. Few people interact with ships compared to other modes of transport (such as cars, trains and aeroplanes) and their levels of safety only reach general awareness during major accidents such as the Costa Concordia grounding in 2012 near the island of Giglio, off the Italian coast. This 'hidden' mode of transport combined with highly litigious environments, where governments will often imprison seafarers after incidents, obscures our understanding of human performance challenges in this domain. As such, the nature of shiphandling should be of interest to human factors researchers. A ship can be more than 300 meters long, weighing more than 250,000 tonnes, operating in dynamic oceans with complex hydrodynamic effects. The complexity of the interaction between humans and these large vessels supports the need for high quality research into perception and situation awareness in this important industry.

**Situation Awareness Defined**

Situation awareness (SA) is commonly defined as 'the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future' (M. R. Endsley, 1988). In Smith and Hancock's Perceptual Cycle Model (1995), situation awareness is viewed as 'a generative process of knowledge creation and informed action' (Salmon et al., 2008). The model is based upon Neisser's Perceptual Cycle (1976), which describes how individuals interact with the world and how they use schemata to modify their interactions with the world. This dynamic approach to achieve SA (Uhlarik & Comerford, 2002) consists of three elements:

- The object (i.e. the available information in the external environment).
- The schemata (i.e. the internally structured knowledge and expectations or goals that are developed through training and experience, and stored in long term memory when not in use).
- The exploration (i.e. a search of the environment by the observer).

Internally held schemata (knowledge) direct a person's exploration of the environment, to facilitate anticipation of situational events by directing one's attention to cues in the environment and their eventual course of action. The person then carries out checks to confirm that the evolving situation conforms to their expectations (K. Smith & Hancock, 1995). The outcome of such exploration may modify the original schemata, which in turn directs further exploration. This process continues in a cyclical manner, obtaining SA as the dynamic result of the interaction of the person with the world; any unexpected event prompts further search and exploration and in turn modifies the existing individual mental model (N. Stanton et al., 2013).

In recent years the discussion around SA was broaden by the introduction of the concept of Distributed Situation Awareness (DSA) (N. A. Stanton, 2016). The difference between Endsley's model (M. R. Endsley, 1988) and this most recent approach is that, in the DSA, a socio-technical system is the unit of analysis, whereas in the three-level SA the individual mind is the unit of analysis. DSA is concerned with the transactions between "agents" (regardless if humans or devices) and the physical structure of the environment in a socio-technical systems (N. A. Stanton, Salmon, Walker, & Jenkins, 2010).

For the purposes of this paper, though, the focus remained on the individuals and on the Perceptual Cycle Model. Our study did not necessarily put the accent on other elements that could have been defined as additional agents in a "socio-technical" system. Very differently from an airplane cockpit, a ship bridge does not provide a standardised instrumentation, nor do marine pilots generally operate in conjunction with co-pilots. Certainly this is not typical in Australian pilotage regimes, and no co-pilots were involved in our experiment. Having said this, the marine pilot is generally not alone on the bridge of a ship. They are typically accompanied by a ship's officer (the Captain or another officer) and a helmsman (who steers the ship at the pilot's direction). For the purposes of this study both of those roles absolutely followed the directions of the pilot because the pilot had the 'conduct' (or control) of the ship.

**Using Eye-Tracking to Inform the Understanding of Situation Awareness**

In the previous paragraph we described how SA is developed through an active search of the environment (Exploration). Understanding how such exploration is carried out, helps us to better appreciate how SA is obtained and maintained. Referring to the previously introduced Perceptual Cycle Model (PCM), SA would derive from the comparison (through the Exploration) between actual representation of the surrounding environment (Object) and the operator's owned knowledge, expectations and goals (Schemata). Perceived discrepancies between Object and Schemata would drive corrective actions. If perceptions were wrong (or partial), the achieved SA would be incorrect, and so would be the deriving corrective actions (or the absence of).

That is why eye-tracking can be used to improve our understanding of situation awareness. In particular, determination of gaze position within the natural environment through the use of eye-tracking, provides important information regarding where overt visual attention is directed. Knowing where operators' attention is focused can help us to better understand what information is used to build and maintain their SA and achieve their goals. Differences in outcomes could be then related to differences in how or what in the object was observed by operators. Ideally, outcomes could be improved correcting or guiding operators to choose and evaluate more relevant and more reliable sources of information. A classic experimental design, that is chosen to identify causes behind differences in outcomes, is the comparison of expert operators against novices. Studying the differences in behaviour of these two groups during a task, can help to understand why outcomes vary and how novices could be supported to reach similar outcomes as experts (Di Stasi, Contreras, Cándido, Cañas, & Catena, 2011; Falkmer & Gregersen, 2005; Underwood et al., 2002).

Eye trackers have been also employed to analyse gaze behaviour in specific and very specialised contexts and conditions, not necessarily focusing on the juxtaposition of two groups. An example could be found in a study conducted on twenty experienced airline pilots to investigate monitoring strategies on automated flight decks. Using eye trackers, the study was able to identify and explain some shortcomings in pilots' automation monitoring strategies (Fisher, Pradhan, Pollatsek, & Knodler Jr, 2007). In general, a large body of work has been conducted in the airline industry, studying the gaze behaviour of pilots in cockpits (Sarter, Mumaw, & Wickens, 2007; Weibel, Fouse, Emmenegger, Kimmich, & Hutchins, 2012). Other studies in the transport industry have considered glance differentiation and distribution while executing in-vehicle tasks (Victor, Harbluk, & Engström, 2005) or drivers with specific visual impairments (Lee, Black, Lacherez, & Wood, 2016).

Moreover, Eye trackers have been extensively utilised for assessing and training in several other fields, such as surgery and medical specialties, human-computer interaction or sports, as summarised in recent reviews (Gegenfurtner, Lehtinen, & Säljö, 2011; Hermens, Flin, & Ahmed, 2013; Rosch & Vogel-Walcutt, 2013; Tien et al., 2014).

The 'hidden' nature of maritime transport means that few gaze behaviour studies have been conducted in the shipping industry. One study compared the gaze behaviour of 16 experienced and novice speed boat drivers and their use of navigational aids (Forsman, Sjors, Dahlman, Falkmer, & Lee, 2012). This study found that novice drivers spent more time looking at proximal objects, such as instrumentation within the cockpit, especially at high speeds, while the experienced drivers looked more at objects further away from the boat. In a simulated maritime study, significant differences in gaze behaviours were observed between novice shiphandlers and experienced captains when overtaking and bypassing a ship in a narrow canal (Muczyński, Gucma, Bilewski, & Zalewski, 2013), where experienced captains focused mainly on observation of the relative positions of the ships and distance between them, and less time monitoring their own ship conning equipment compared to the novice shiphandlers. Another simulator study investigated differences in gaze distribution between a small sample of 4 students and 2 experienced navigators while conducting a ship in constrained waters (M. Lützhöft & Dukic, 2007), and showed that experienced navigators made less glances per minute and generally followed a more organised scanning pattern than the students. A recent study compared eye tracking data collected on board of a littoral combat ship at sea and in a bridge simulator (Hareide, Ostnes, & Mjelde, 2016). In this research gaze behaviour was similar in the two environments with respect to gaze dwelling time on specific areas of interests, suggesting the validity and usefulness of simulators to address these types of research questions.

Overall, limited research has been carried out, specifically assessing Marine Pilots in simulator environments. In the context of the maritime industry, Marine Pilots are seafarers that are specifically trained and certified to manoeuvre vessels within well-defined critical coastal and port waters. They embark a ship outside port waters and then work with the bridge team to navigate the ship to berth. A Canadian report (CMPA, 2017) highlights the importance of Marine Pilots, as one of the most effective measures to mitigate accidents. The report identified that piloted ships have a 44 times lower risk of accidents compared to those that are not piloted (from 0.094 to 0.0021 probability of accident per vessel).

**This study in the context of PCM and the previous research.**
As explained in the previous paragraphs, gaze can be profitably used in the context of PCM to analyse the Exploration, the fundamental link between the subjects' Schemata and the surrounding environment, the Object. As suggested by previous studies, differences in subjects Schemata (for example differences in knowledge / experience content between experts and novices) are expected to elicit different exploration behaviours. The current study explores this phenomenon.

To achieve this, we could not rely on a simple comparison between novice and experts, since all our subjects were experienced pilots. We chose instead to control another experimental factor: the experimental manoeuvres. The experimental manoeuvres, with their different levels of difficulty and their different phases, were used to elicit different pilots' schemata. As described in a previous publication (Luca Orlandi et al., 2015) pilots' schemata were collected, through a face to face structured interview, in the form of Detailed Manoeuvring Plans (DMP). Those DMP were collected in a tabular form, before performing manoeuvres in the simulator. Those tables can be seen as a more detailed version of the routine passage plan normally discussed by pilots and ship masters before a ships enters into a port (Wild & Constable, 2013). The DMP obtained, included also navigational charts with sketches depicting the expected ship movement and highlighting various elements of interest. A DMP was a record of pilots' expectations, goals and strategies to achieve specific goals throughout the manoeuvres. In the context of PCM those plans can be considered pilots' Schemata. For the purpose of this study, the schemata so obtained are summarised in Table 3 presented in the discussion.

For the gaze analysis carried out in this research (and differently from previous studies), we decided to focus not simply on single objects or areas of interests, but on specific combinations or sequences of objects as they were targeted by gaze. With this goal in mind, a series of labels, adopted in the video coding, were defined to identify the specific objects targeted by the gaze (instruments such as radar or log, navigational aids such as buoys, etc..). In addition, specific sequences of labels (see table 2), expected to identify goal orientated explorative behaviours, were investigated, while pilots completed the experimental manoeuvres. Those identified sequences of gaze labels allowed the creation of a list of dependent variables. Those variables, measuring the number of times per minute those sequences were performed by each subject, were then used for statistical analysis.

The study's hypothesis was that different types of manoeuvres and different phases of these manoeuvres, eliciting different schemata, would create changes in combinations of gaze responses. If specific sequences can be clearly associated with the research of a certain type of information (exploration), this can indirectly inform about the schemata driving the search. Relating those schemata (internal knowledge) with shiphandling outcomes, could help to better understand pilots' expertise, with important implications for safety, training and assessment (Luca Orlandi et al., 2014).

## Methodology

**Participants**
Participants included 10 marine pilots (mean age = 51.8 ± 5.9 years), employed by an Australian pilot company. The marine pilots were all males and in good health, as required by national professional medical standards set by the Australian Maritime Safety Authority (AMSA, 2010). According to those standards, pilots are required to pass a biannual (or yearly, if over 55) full medical check. Participants on average had more than ten years of previous experience as qualified pilots (mean experience = 10.6 ± 7.8 years). There was no significant difference in age ($p = 0.79$) and years of service ($p = 0.89$) of the participants and the rest of the pilots working for the same company. The research complied with the ethical standards of the University of Tasmania.

**Experimental Setting**
The Maritime Safety Queensland Simulator located in Brisbane was used (Smartship® Simulator www.smartshipaustralia.com.au). This 'Full Mission Bridge' simulator runs Force Technology® software (www.forcetechnology.com) and is classified as Class A (NAV) according to the standards issued by the classification society Det Norske Veritas Germanischer Lloyd (DNV_GL_AS, 2014). This simulator is capable of replicating a total shipboard bridge operation, including advanced

manoeuvring in restricted waterways. Before the experimental manoeuvres, pilots performed a simple mooring with a vessel different from those used in the experimental runs, to gain practice and become familiar with the simulator environment.

**Apparatus**

Eye movements were recorded using a pair of light-weight eye tracking goggles (ASL® Mobile Eye XG® - www.asleyetracking.com). This eye-tracker system comprises two cameras each sampling at 30 Hz: a forward facing scene camera and an eye camera, capturing the infrared corneal reflection and pupil position of the right eye. A calibration procedure, which determines where gaze is located within the scene, was performed at the beginning of each manoeuvre. The full calibration is obtained asking to the subject to fix a specific element present in their field of view. Using a laptop computer, wirelessly connected to the eye trackers, it is possible to click with a mouse on the same object visible on the pc screen and aimed by the observer, correcting the initial offset that the crosshair may show from the designated object. This operation is repeated several times for different objects located in different areas of the tracker's field of view, until the crosshair identifies precisely the gaze location. This procedure was carried out at the beginning of each experimental manoeuvre.

**Coding Procedure**

Anvil (Kipp, 2010) is a free video annotation tool, that offers multi-layered annotation based on a user defined list of labels. To code, the user can place labels on multiple tracks in time-alignment with the video analysed. A series of labels was adopted to identify each object targeted by the gaze, according to the position of the crosshair on the video (equipment on the bridge, rudder controls, engine levers etc). When the gaze was directed to objects outside of the bridge windows (port features and / or navigational aids), in the coding was also included the relative bearing of where the gaze was directed. The relative bearing was measured clockwise from the bow of the vessel, on a 360 degrees angle at intervals of 30 degrees. If, for example, the pilot was looking at an object at the beam of the vessel, the label identified the type of object, followed by a number which specified if the object was, for example, on the left (270) or on the right side (090) of the ship. To quantify any other direction, the frames of the bridge windows could be easily used as a reference. Using programming scripts, it was possible to process the files created by the video coding application and identify within those files how often specific gaze sequences were occurring.

All the video coding identified here was coded by the same coder, the lead author. To verify the reliability of the coding procedure, one run among the 40 completed, was entirely recoded by the same coder. The second video coding was then compared against the first, obtaining a total of 478 coding instances, of which: 354 were a match; 75 were codes present in the recoding but not matching the original; 49 were codes present in the original but not matching the recoding. A Cohen's Kappa (1960) value of 0.73 was obtained. In addition, using a probability of 95%, a confidence interval was calculated (McHugh, 2012), obtaining Kappa values included between 0.69 and 0.77. According to Landis and Koch (1977), K values included in the obtained C.I. indicate a substantial agreement.

Inter-rater reliability analysis was performed using a naive coder who had not participated in the collection of the data, and their coding was compared to the coding of the lead author. This second coder had no marine pilotage expertise, and is herein called the 'naïve coder'. The naïve coder was provided with a 90 minute training session on the coding tool. Subsequent to this, a coding task was performed, where the coder inserted 93 codes during a 'swing' phase of a manoeuvre (see paragraph "Independent Variables"). Comparison with the original coding returned a Cohen's Kappa value of 0.50. This was considered unacceptable and the naïve coder was retrained identifying several errors in their coding activity. Subsequent to this, a further coding task was performed, where the coder inserted 166 codes during a swing phase of a manoeuvre. Comparison with the original coding returned a Cohen's Kappa value of 0.65 (194 coding instances in total; 129 were a match; 37 were codes present in the naïve coding but not matching the original; 28 were codes present in the original but not matching the naïve coding). In addition, using a probability of 95%, a confidence interval was calculated, obtaining Kappa values included between 0.58 and 0.72. This, and the complete recoding of an entire sequence by the main coder suggest that the coding scheme was reliable.

## Experimental Design

To investigate pilots' perceptual cycle (K. Smith & Hancock, 1995), four different berthing manoeuvres were set as experimental conditions (Object). Each manoeuvre included the whole process necessary to transfer the ship from a defined initial position to a berth within constrained port waters, with the use of their own and/or external means of propulsion (i.e. tug boats to assist, when allowed). The manoeuvres were presented to pilots before being performed in the simulator, since every participant was required to provide a detailed manoeuvring plan (DMP). All DMPs were collected and served as Schemata for the purpose of this study. The detailed explanation of how the DMPs were collected and analysed, can be found in a dedicated previous study (Luca Orlandi et al., 2015). The Schemata obtained from the DMPs are summarised in Table 3.

## Independent Variables

Four manoeuvres were conducted by each pilot, the order of which was randomised between participants, to minimise possible learning effects. Two ports were selected for the research: the pilots' homeport, where they normally work, and a foreign virtual port, developed in the simulator software. Spatial constraints and port dimensions were purposely maintained to be as similar as possible between the two ports, modifying the virtual port in order to match the homeport dimensions. The virtual port was used to avoid any possibility of a learning effect associated with previous manoeuvring experience that the participants may have had and to provide support for methodology reliability. The two ports were not considered as an independent variable, but the data collected from the two different ports was considered as a repeated measure within the subjects, for each level of difficulty.

The 'Difficulty', defined by 2 levels (Easy and Hard), was the first independent variable of this study. The two levels of 'difficulty' were based on a range of parameters manipulated in the simulator, as outlined in Table 1. The easy manoeuvres were comparable to those of routine operations, while hard manoeuvres provided pilots with a scenario which marginally exceeded the safety limits established in the pilots' homeport, without losing construct validity.

**Table 1. Levels of Difficulty – Adopted in both Ports.**

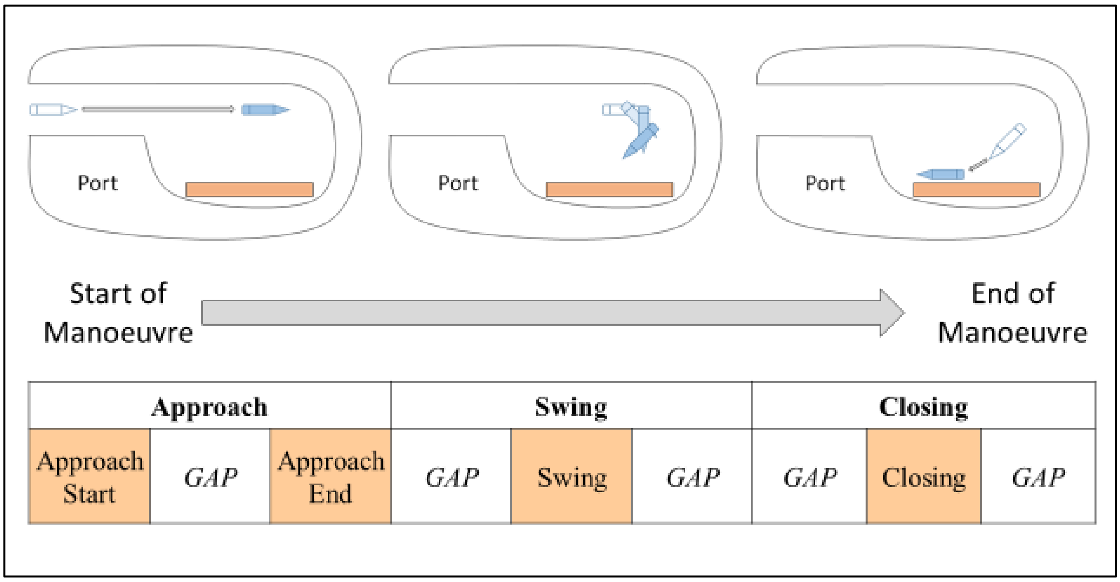
| | **Pier - Spatial constraints** | **Environmental conditions and forces** | **Vessel characteristics** | **Tugs** | **Interactions with traffic** | **VTS Comms (1)** |
|---|---|---|---|---|---|---|
| **Easy Level** | Big Swing Basin<br><br>(3 times Vessel LOA (4)) | Current: 0.7 Knt<br><br>Wind: 15 Knt<br><br>Good Visibility | Single Controllable Pitch Propeller (2)<br><br>Bow Thruster (3) | None | 1 Interacting but not Interfering vessel | General Info<br><br>No reporting Points |
| **Hard Level** | Small Swing Basin (5)<br><br>(1.5 times Vessel LOA) | Current: 2 Knt<br><br>Wind: 25 Knt<br><br>Poor to no Visibility - Heavy Rain | Single Fixed Pitch Propeller<br><br>No Thrusters | As required by Pilot | 1 Interacting<br>1 Interfering vessels | General Info and Traffic Advice<br><br>Reporting Points |
| **Notes** | (1) Vessel Traffic Management station present in a port and managing ships via radio communications;<br>(2) Propeller capable to change the water thrust direction changing the angle of the blades instead of direction of rotation;<br>(3) A thruster is a propeller positioned perpendicular to the ship keel axis. Placed on the bow or on the stern, induces transversal / angular motion;<br>(4) Length Over All, maximum length of a vessel;<br>(5) Wider area, within constrained waters, where ships have sufficient room to rotate and revert their direction. | | | | | |

As anticipated, for each level of difficulty, two manoeuvres were conducted (and considered as repeated measures): one in pilot's homeport, and the other one in the foreign port. Each manoeuvre required pilots to complete a mooring using the side of the ship opposite to the berth position on commencement of the manoeuvre. This implied that the ship had to swing (rotate 180°) before she could be moored. Each manoeuvre therefore had to be developed through three main phases: the ‘approach’ (from the initial position until the start of the swing), the ‘swing’ (from the start of the swing until the rotation was completed and the ship stabilised), and the ‘closing’ (from the end of the swing until a defined distance from the berth). Figure 1 provides a schematic detailing the three phases identified in each manoeuvre.

Due to the time necessary to complete manual coding of the whole duration of the manoeuvres (each 10 minutes of a video clip would require approximately 3 hours of manual coding), a sampling strategy was applied. Coding windows were chosen within the three phases of the manoeuvres. A total duration of 20 minutes of video coding was obtained for each manoeuvre, considering four different sections, each of at least five minutes duration (unless the duration of that specific manoeuvring phase was less). The average duration was 32 minutes (± 1.5 minutes) for the easy manoeuvres and 1 hour and 13 minutes (± 12.5 minutes) for the hard ones, so the coding windows covered ~64% of the easy and ~27% of the difficult manoeuvres. As shown in Figure 1, the coding windows’ location in each manoeuvre was chosen based on specific criteria. In the Approach phase, two coding windows were selected, one at the beginning (Approach_Start) and one at the end (Approach_End) of the phase. Two coding windows were located in this phase (compared to only one for each of the other phases) since the Approach was generally longer than the other two phases (especially in the hard manoeuvres). For the swing and the closing phases, the coding window was located in the middle. Time gaps between coding windows varied according to the total duration of the different phases.

Those four ‘coding windows’, were adopted as the second independent variable of this study.

**Figure 1. Schematic of a manoeuvre with its three phases (approach, swing and closing) and the coding windows from which data were analysed.**




### Dependent Variables

In this study, three specific sequences were adopted as dependent variables: Position Check, Direction Check and Speed Check. The list of all the sequences defined as dependent variables is reported in Table 2.

**Table 2. Dependent Variables - Sequences of gaze labels.**

| Sequence Name | Sequence Description (Targeted gaze labels) | Sequence Explanation |
|---|---|---|
| Position Check | A first glance to an external object (i.e. buoy) in a certain relative direction (90 ° sector) and then, within 3 seconds, a second glance to an external object in a direction perpendicular to the previous. | Sequentially comparing objects at approximately 90 degrees from each other, facilitates pilots understanding of where they are located in the manoeuvring space. |
| Direction Check | Shift of gaze from the bow to an external object direction within 30° off the bow. | Moving gaze between an external object within 30 degrees centred on the bow and another object in the same sector and / or a heading instrument (gyro repeater), allows pilots to perceive and monitor vessel direction. |
| Speed Check | A glance to an external object at the beam of the vessel (relative bearing 090° or 270°, followed or anticipated by a glance on a speed sensor (LOG or GPS SOG indicator) | Moving gaze alternatively from an external object at the beam of the vessel to a speed sensor (i.e. Log) allows pilots to evaluate vessel's speed. |

In Figure 2 it is shown an example of video coding resulting in a sequence defined as Speed Check. In the video clip, between instants 30:54 and 30:57 (duration of 3 seconds) it was possible to observe the gaze (crosshair) fixing an object on the left side of the vessel (bearing 270 degrees). Between instants 30:58 and 31:02 (duration of 4 seconds) the gaze was monitoring a position on a bridge screen were it was reported the vessel's speed. The two codes in a sequence were considered as a single Speed Check.

**Figure 2. Example of video coding resulting in a Speed Check**


For each of the previously described coding windows, the frequency (checks per minute) of the dependent variables defined in Table 2, was recorded every 15 seconds. The maximum frequency score for each dependent variable recorded in each coding window, was chosen for statistical analysis.
For each dependant variable, a data table with a total of 160 cases (10 subjects X 4 manoeuvres (2 Easy - 2 Hard) X 4 coding windows) was obtained and analysed.

**Missing Data**
Across the 40 manoeuvres, four crashes were recorded. A crash was defined as an impact or grounding that required the termination of the simulated manoeuvre and data collection. All crashes occurred during the hard manoeuvres. More specifically, two crashes occurred during the approach phase, which meant that no data could be collected for the following swing and closing phases. Two crashes occurred during the swing, which meant that no data could be collected in the following closing phase. Impacts were classified as a contact of the vessel with another ship or port infrastructures, but did not impede the continuation of the manoeuvre. Cases with missing values due to clashes were included in the analysis, and the missing values were left missing (empty) in the data table. In Appendix A the letter "M" was inserted in all those cases where data was missing.

**Statistical Methods**
To examine whether the gaze sequences (dependent variables) varied during the execution of the manoeuvres, a Generalised Estimating Equations analysis was performed using the statistical package IBM SPSS (IBM_Corp., 2010). Main effects and interactions were assessed, and p-values less than 0.05 were considered significant. The two factors included in the analysis were 'Difficulty' (Easy and Hard) and 'Coding Window' (Approach_Start, Approach_End, Swing and Closing) for each of the dependent variables (Position, Direction and Speed Checks). The data collected from the two different ports was considered as a repeated measure within the subjects, for each level of difficulty. When the factor 'Coding Window' showed a significant main effect, a post hoc comparison was carried out adopting a Bonferroni adjustment. The effect size reported, was obtained as the difference between the considered estimated means divided by the relevant dependent variable standard deviation. Different models were compared using the Quasi Likelihood under Independence Model Criterion, confirming that the best Goodness of Fit was obtained choosing the 'Negative Binomial with log link' distribution.

## Results

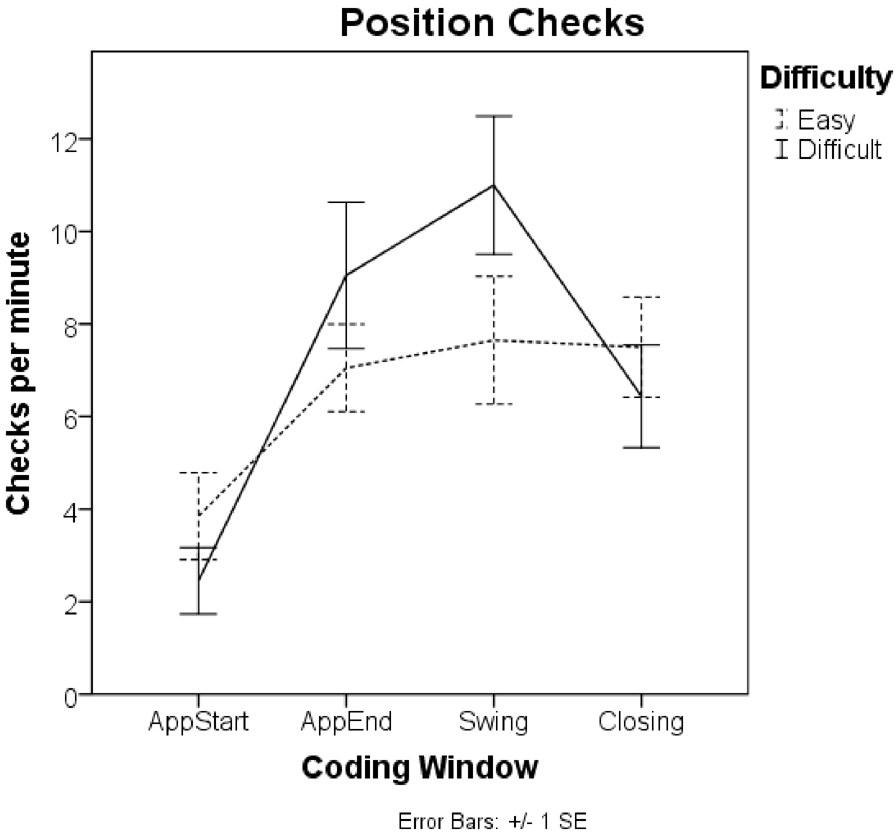
### Position Check

As shown in Figure 2, the frequency of Position Checks significantly changed across the coding windows (Wald $\chi^2(3) = 83.45$, $p < 0.001$). A significant increase ($p < 0.001$) with a large effect size ($d = 0.86$) was recorded from the Approach_Start (3.1 checks per minute) to the Approach_End (8.0 cpm). The checks significantly increased ($p < 0.001$) between the Approach_Start (3.1 cpm) and the Swing (9.2 cpm) with a large effect Size ($d = 1.07$) and between Approach_Start (3.1 cpm) and the Closing (6.9 cpm) ($p = 0.02$) with a medium effect size ($d = 0.68$). There was no main effect on the factor Difficulty (Wald $\chi^2(1) = 0$, $p = 0.98$), however, there was a significant interaction between Difficulty and Coding Window (Wald $\chi^2(3) = 19.88$, $p < 0.001$).

In the simple effects analysis, in the 'easy' condition, position checks significantly differed across some of the coding windows (Wald $\chi^2(3) = 15.13$, $p < 0.01$). The easy manoeuvres showed that the position checks increased from the Approach_Start (3.8 cpm) to the other coding windows (even though not always significantly). More specifically, it was recorded a significant increase ($p < 0.01$) between Approach_Start (3.8 cpm) and Approach_End (7.0 cpm) with a medium effect size ($d = 0.63$). The increase between Approach_Start (3.8 cpm) and the Swing (7.6 cpm) was not significant ($p = 0.07$). A significant increase ($p = 0.05$) was recorded between Approach_Start (3.8 cpm) and the Closing (7.5 cpm) with a medium effect size ($d = 0.72$).

In the simple effects analysis, in the 'difficult' condition, position checks significantly differed across some of the coding windows (Wald $\chi^2(3) = 125.49$, $p < 0.001$). The position checks increased from the window Approach_Start (2.5 cpm) to the other coding windows (even though not always significantly). More specifically, it was recorded a significant increase ($p < 0.001$) between Approach_Start (2.5 cpm) and Approach_End (9.0 cpm) with a large effect size ($d = 1.04$). A significant increase ($p < 0.001$) was recorded between Approach_Start (2.5 cpm) and the Swing (11 cpm) with a large effect size ($d = 1.35$). The increase between Approach_Start (2.5 cpm) and the Closing (6.4 cpm) was not significant ($p = 0.07$).

**Figure 3. Group mean data for the number of times per minute participants compared objects at 90 degrees from one another as a function of the manoeuvring coding window.**




**Direction Check**

As shown in Figure 4, the frequency of Direction Checks significantly changed across the coding windows (Wald $\chi^2(3) = 20.56$, $p < 0.001$). A significant decrease ($p = 0.03$) with a medium effect size ($d = 0.51$) was recorded from the Approach_Start (11.0 cpm) to the Approach_End (7.7 cpm). The checks significantly decreased ($p < 0.01$) between the Approach_Start (11.0 cpm) and the Swing (5.0 cpm) with a large effect size ($d = 0.91$) and between Approach_Start (11.0 cpm) and the Closing (6.2 cpm) ($p < 0.01$) with a medium effect size ($d = 0.74$). The decrease between the Approach_End (7.7 cpm) and the Swing (5.0 cpm) was also significant ($p < 0.001$) with a small effect size ($d = 0.40$). There was no main effect on the factor Difficulty (Wald $\chi^2(1) = 3.75$, $p = 0.053$), however, there was a significant interaction between Difficulty and Coding Window (Wald $\chi^2(3) = 10.68$, $p = 0.01$).

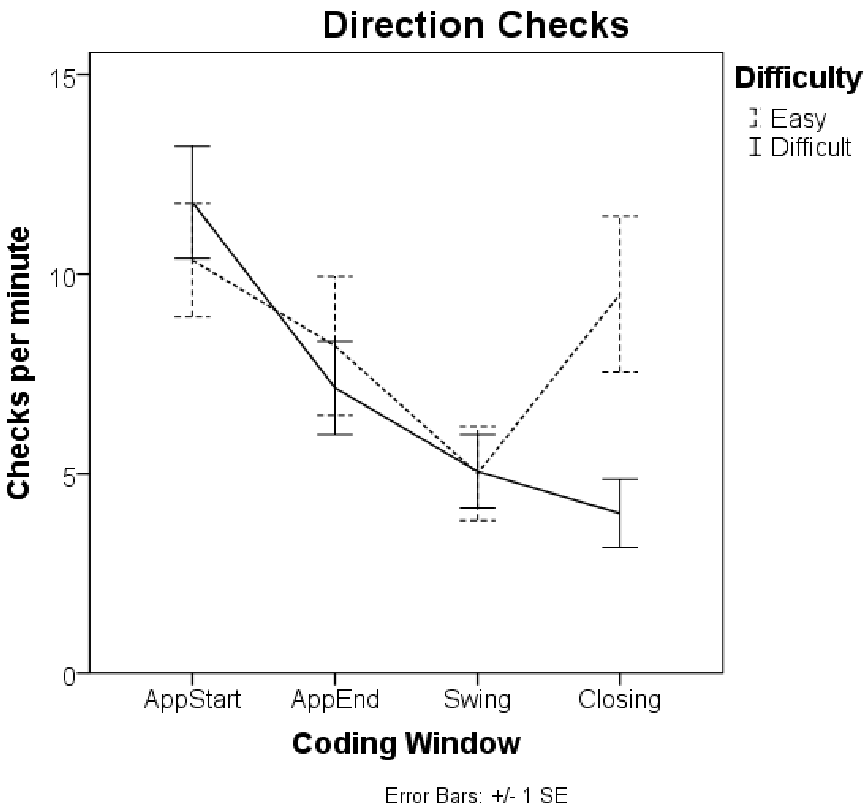
In the simple effects analysis, in the 'easy' condition, direction checks significantly differed across some of the coding windows (Wald $\chi^2(3) = 11.45$, $p = 0.01$). The easy manoeuvres showed that the direction checks decreased from the Approach_Start (10.3 cpm) to all the other coding windows (not always significantly). More specifically, the decrease between Approach_Start (10.3 cpm) and the Approach_End (8.2 cpm) was not significant ($p = 0.16$). The decrease between Approach_Start (10.3 cpm) and the Swing (5.0 cpm) was significant ($p < 0.01$) with a medium effect size ($d = 0.73$). The decrease between Approach_Start (10.3 cpm) and the Closing (9.5 cpm) was not significant ($p = 1.0$).

In the simple effects analysis, in the 'difficult' condition, direction checks significantly differed across some of the coding windows (Wald $\chi^2(3) = 22.94$, $p < 0.001$). The difficult manoeuvres showed a negative trend in the direction checks from the beginning to the end of the manoeuvres (not always significantly). More specifically, the decrease between Approach_Start (11.8 cpm) and Approach_End

(7.1 cpm) was not significant ($p = 0.07$). A significant decrease ($p < 0.01$) was recorded between the Approach_Start (11.8 cpm) and the Swing (5.0 cpm) with a large effect size ($d = 1.18$). A significant decrease ($p < 0.001$) was recorded between the Approach_Start (11.8 cpm) and the Closing (4.0 cpm) with a large effect size ($d = 1.37$).

The interaction can be noticed in the trend shown in the Closing window, which is increasing for the 'easy' manoeuvres (9.5 cpm) and decreasing for the 'difficult' ones (4.0 cpm).

**Figure 4. Group mean data for the number of times per minute participants shifted gaze from the bow to a direction within 30° off the bow as a function of the manoeuvring coding window.**



**Speed Check**
As shown in Figure 5, the frequency of Speed Checks significantly changed across the coding windows (Wald $\chi^2(3) = 103.02$, $p < 0.001$). A significant increase ($p < 0.001$) with a large effect size ($d = 0.83$) was recorded between the Approach_Start (0.5 cpm) to the Approach_End (2.8 cpm). The checks significantly increased ($p < 0.001$) between the Approach_Start (0.5 cpm) and the Swing (3.0 cpm) with a large effect size ($d = 0.93$) and between Approach_Start (0.5 cpm) and the Closing (4.4 cpm) ($p < 0.001$) with a very large effect size ($d = 1.42$). The increase between the Approach_End (2.8 cpm) and the Closing (4.4 cpm) was also significant ($p = 0.02$) with a medium effect size ($d = 0.59$).

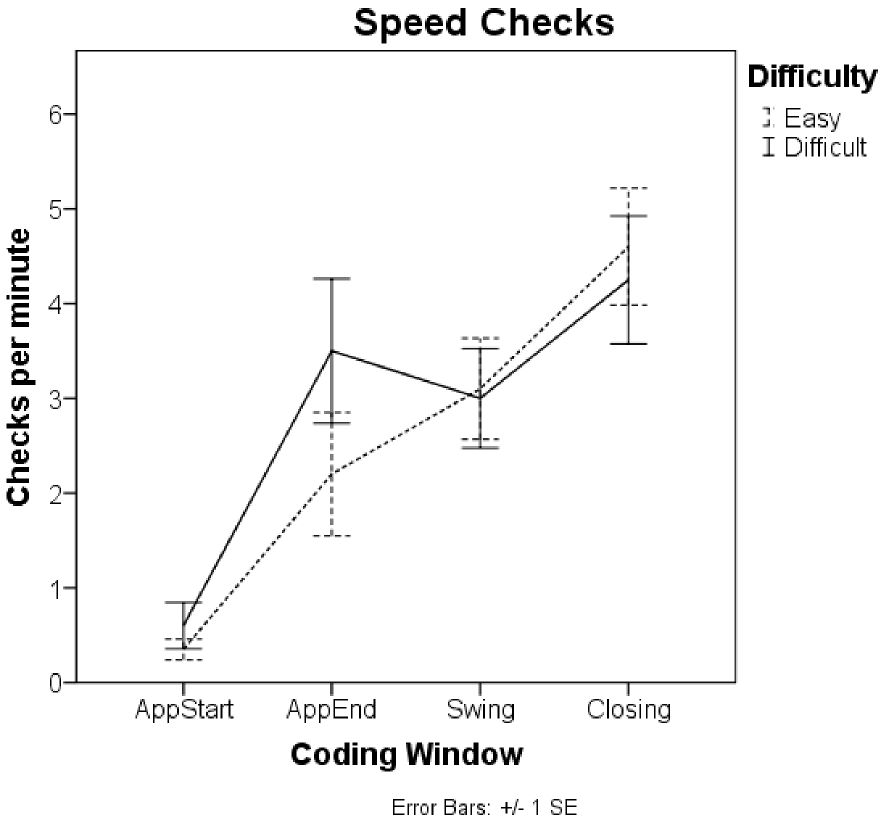
There was a main effect on the factor Difficulty (Wald $\chi^2(1) = 4.36$, $p = 0.04$), with a significant increase of speed checks between the Easy (1.8 cpm) and the Difficult (2.3 cpm) condition, but a very small effect size ($d = 0.16$),

It was recorded also a significant interaction between Difficulty and Coding Window (Wald $\chi^2(3) = 14.26$, $p < 0.01$).
In the simple effects analysis, in the 'easy' condition, speed checks significantly differed across all the coding windows (Wald $\chi^2(3) = 68.69$, $p < 0.001$). The easy manoeuvres showed results in the Coding Windows increasing from the Approach_Start (0.3 cpm) to all the other coding windows. More specifically, a significant increase ($p = 0.04$) was recorded between the Approach_Start (0.3 cpm) and the Approach_End (2.2 cpm) with a medium effect size ($d = 0.67$). The increase between the Approach_Start (0.3 cpm) and the Swing (5.0 cpm) was significant ($p < 0.001$) with a large effect size ($d = 0.99$). The increase between Approach_Start (0.3 cpm) and the Closing (4.6 cpm) was significant ($p < 0.001$) with a very large effect size ($d = 0.99$). It was significant also ($p = 0.02$) the increase between the Approach_End (2.2 cpm) and the Closing (4.6 cpm) with a large effect size ($d = 0.86$), and the increase ($p < 0.02$) between the Swing (3.1 cpm) and the Closing (4.6 cpm) with a medium effect size ($d = 0.54$).

In the simple effects analysis, in the 'difficult' condition, speed checks significantly differed across all the coding windows (Wald $\chi^2(3) = 34.86$, $p < 0.001$). In the difficult manoeuvres the direction checks increased from the Approach_Start (0.6 cpm) to all the other coding windows. More specifically, a significant increase ($p < 0.001$) was recorded between the Approach_Start (0.6 cpm) and the Approach_End (3.5 cpm) with a large effect size ($d = 1.03$). The increase between the Approach_Start (0.6 cpm) and the Swing (3.0 cpm) was significant ($p < 0.001$) with a large effect size ($d = 0.85$). The increase between Approach_Start (0.6 cpm) and the Closing (4.2 cpm) was significant ($p < 0.001$) with a very large effect size ($d = 1.30$). Differently from what encountered in the analysis of the easy manoeuvres, in the difficult manoeuvres no other pairwise comparison was significant.

The interaction can be noticed in the higher number of checks performed in the Approach_End window during difficult manoeuvres (3.5 cpm) compared to easy ones (2.2 cpm).

**Figure 5. Group mean data for the number of times per minute participants shifted gaze from the beam of the vessel to a speed sensor as a function of the manoeuvring coding window.**




**Discussion**

Referring to Smith and Hancock's Perceptual Cycle model of situation awareness (K. Smith & Hancock, 1995), this study demonstrated that pilots' scanning ('the exploration') of the environment ('the object') was dependent upon different shiphandling conditions and pilots' priorities (elicited as different 'Schemata'). Through the use of eye tracking in a simulator, this study was able to map changes in gaze behaviour of marine pilots, providing an example of how to apply quantitative analysis in this domain. Table 3 describes more in details marine pilotage situational awareness according to Smith and Hancock's Perceptual Cycle model (1995). In Table 3 we reported in columns the different coding windows into which the manoeuvres were divided. For each column (each coding window) different elements of our discussion were further assigned to different rows, dedicated to the three constituents of the PCM, the schemata, the object and the exploration. More in details: The first row (Description) reports the context, the manoeuvring conditions at the time of the coding window. According to the PCM nomenclature, this would be the 'Object'. The second row (Shiphandling Priority or Goal) explains what was the most important result that pilots needed to achieve during that coding window. Such goal, being held by pilots as internal knowledge, was considered the 'Schemata'. The third row (Relevant pilot knowledge), also was considered as 'Schemata'. This row reports the elements that specifically depict those structures of internal knowledge and previous experience held in the long term memory (D. A. Lieberman, 2011), that can be dynamically recalled (J. R. Wilson & Rutherford, 1989) in the working memory as mental models (Johnson-Laird, 1983), and can be directly involved in the achievement of "recognition-primed decisions" (G. A. Klein, Calderwood, & MacGregor, 1989).

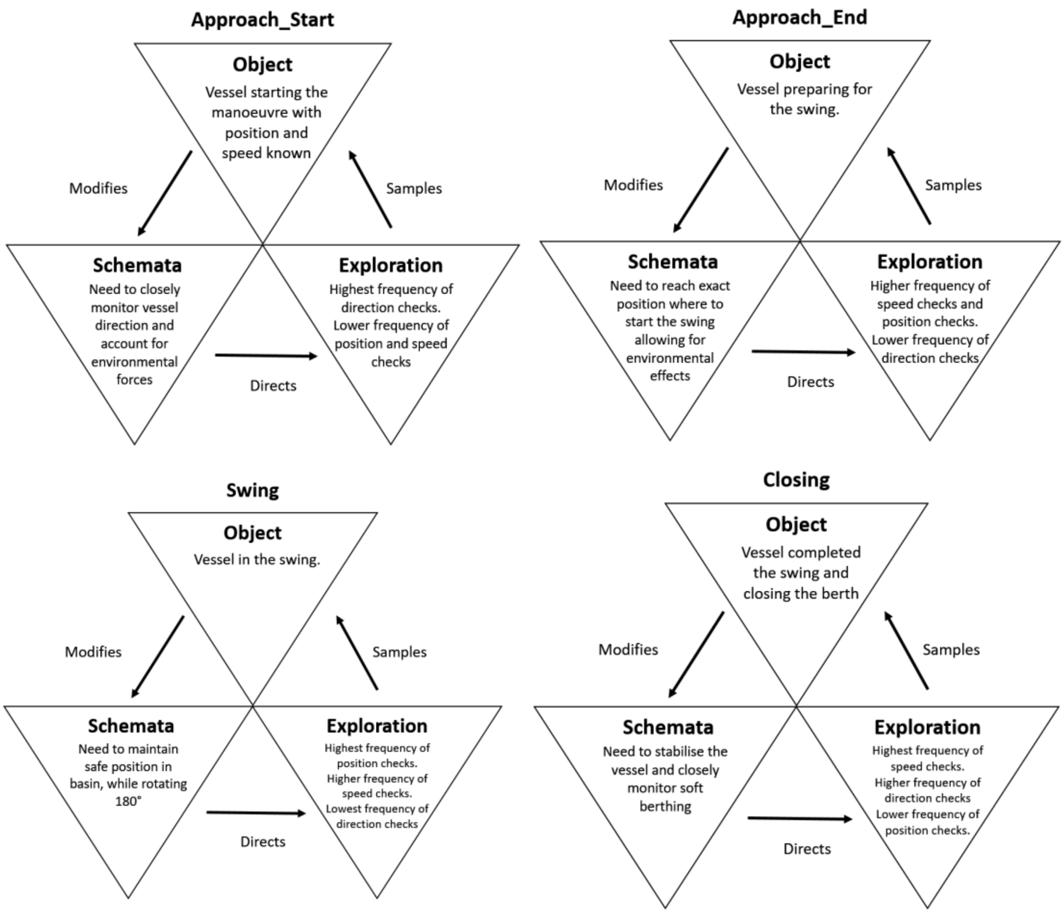
A previously published study (Luca Orlandi et al., 2015) identified that the description of those goals and knowledge structures was directly obtained from pilots through a face to face interview. These interviews, similarly to the approach used by Plant and Stanton (2013), adopted a semi structured approach, to categorize Pilots' explanations about the manoeuvres they were planning to conduct in the simulator. Differently from Plant and Stanton, though, during those interviews, the collection of information went well beyond a simple qualitative level (Boyatzis, 1998). From each pilot, for each experimental manoeuvre, a detailed manoeuvring plan (DMP) was obtained. Such detailed manoeuvring plan not only collected a brief description of pilots' considerations and goals, but aimed to numerically capture measurable quantities, such as the expected use of engine, the expected forces exerted by tugs in assistance, the expected position of the vessel. Pilots had to describe their general understanding of what they thought was going to happen, and also were asked to provide in their detailed manoeuvring plans, numerical quantifications of measurements intimately related to those future expectations and goals. For the purposes of this paper, we will simply summarise those results as the proposed schemata. The fourth row summarises the Gaze Behaviour, which according to the PCM, would underlie the 'Exploration'. This was the particular element of the PCM (and its interdependence with the other two), that was the specific focus of this study.

**Table 3. Perceptual Cycle Model applied to results.**

| Coding Window | Approach_Start | Approach_End | Swing | Closing |
|---|---|---|---|---|
| **Description (Object)** | Start of manoeuvre with vessel in a known position and at a desired speed, directed towards the berth | Vessel approaching the position were the swing is started | Completion of the 180° rotation of the vessel within the swinging basin spatial constrains | Final closing to the berth with suitable angle and speed in order to safely moor the vessel |
| **Shiphandling Priority or Goal (Schemata)** | At the starting of the manoeuvre, the vessel is still relatively distant from the expected area of rotation (swing basin). The priority at this stage is to monitor the direction and manage the speed at which the vessel is proceeding. The aim is to appreciate and counteract the effects of environmental forces on the vessel so to manoeuvre her correctly in order to reach the desired position where to safely start the swing. | At this stage, the speed should have already been consistently reduced in order to reach in a very controlled manner the exact position where the swing is expected to start. The priority is to position the vessel correctly, in order to have enough space to swing it around in the next phase.<br>Another goal is also to correctly reposition the tugs (when available) to have them applying the desired forces. | In this phase the vessel is completing the 180° turn. The rotation is obtained using tugs to apply transversal forces, while the main engine is used to control longitudinal momentum. In addition, the vessel continuously changes its relative angle to the environmental forces, developing new resultants that require to be managed / corrected to maintain the desired safe position. | The vessel, once stabilised her angular momentum at the end of the swing, has to be controlled in angle and speed, during her approach to the berth. Longitudinal and lateral speeds have to be reduced to a minimum before touching the fenders. |
| **Relevant pilot knowledge (Schemata)** | Known heading values to be matched to aim in the right direction.<br>Known safety envelopes (max allowed speed / angle difference from expected heading) to ensure and maintain positive control of vessel.<br>Known hydrodynamic effects of wind and current on the actual direction and speed of the vessel that will have to be accounted for and counteracted to maintain the vessel in a safe position. | Knowledge of desired position to be reached in consideration of future effects that the vessel will experience during the swing. | Knowledge of the effects of environmental forces while the vessel is constantly changing angle. | Knowledge of heading values to be matched to have a suitable angle of approach with the berth.<br>Knowledge of safety envelopes (max allowed speed / angle difference from expected heading) to ensure positive control of vessel during landing.<br>Knowledge of effects of wind and current depending on the actual direction and speed of the vessel. |
| **Gaze Behaviour (Exploration)** | Lower number of Position Checks:<br>Not fundamental at this stage.<br><br>Higher number of Direction Checks:<br>Use of visual cues against the forward mast of the ship, to promptly perceive rotation and verify direction in the channel;<br>Use of dedicated equipment to check heading and rotation.<br><br>Lower number of Speed Checks:<br>Limited use of visual appreciation of vessel's speed looking at objects at the beam. | Higher number of Position Checks:<br>Increase due to the need to stop the vessel exactly in the desired position to safely initiate the swing;<br><br>Decrease in number of Direction Checks:<br>Direction now less relevant due to reduced speed.<br><br>Increased number of Speed Checks:<br>Ship needs to be stopped. The checks are even higher in the difficult manoeuvres. | Highest number of Position Checks:<br>The ship has to maintain a safe position throughout the whole duration of the swing;<br><br>Lowest number of Direction Checks:<br>Direction not relevant since ship is fully rotating.<br><br>Increased number of Speed Checks:<br>Speed needs to closely controlled to avoid undesired movements towards port infrastructures. | Drop in number of Position Checks:<br>The ship is close to the berth.<br><br>Increase in number of Direction Checks:<br>Only in the easy manoeuvres, where the distance from the berth after the swing justifies further checks in direction.<br><br>Highest number of Speed Checks:<br>Speed needs to really closely monitored to avoid damage on ship and berth due to excessive momentum. |

To highlight the cyclical nature of the pilot's perception during the manoeuvre, a schematic translating Table 3 in the constituting elements of the PCM, was created for each coding window in Figure 6. These figures follow the design used by Plant and Stanton (2012), where the PCM was adopted for the analysis of an airline incident.

**Figure 6. PCM Schematics depicting Schemata, Object and Exploration for each Coding Window.**




A detailed discussion, dedicated to the results obtained in each dependent variable, is proposed in the following paragraphs.

**Position Check**

Position Checking is an important scanning behaviour for pilots, providing information on their position, or change of position, within the port. Given the dimensions and inertia involved in manoeuvring large vessels, appreciation of position and motion is obtained through monitoring how objects in the foreground change relative to objects located in the background. Choosing to observe objects along different directions around the vessel allows pilots to 'triangulate' the ship's position. Our study showed that Position Checks, as a main effect, increased from Approach_Start to Approach_End and from Approach_Start to Closing. This can be explained by the fact that when the study's manoeuvres started, the ships were in a known position and proceeding at speed, so for pilots it was more important to monitor the direction the ship was heading to, rather than checking its actual position. Later on in the manoeuvre (even though always important), it became particularly relevant for pilots to be confident about the exact position of the ship. Approach_End is when and where pilots decided to start the swing and ensure that the ship was in a suitable position, allowing sufficient clearance to rotate with enough distance from surrounding obstructions. Figure 3,

demonstrates that in the hard manoeuvres, a significant reduction in scanning frequency occurred between the Swing and the Closing. This suggests that during the final segment of the manoeuvre, pilots do not have to repeat such behaviour as often as in previous phases, probably due to the close vicinity of the berth.
In Figure 3 it is also possible to observe how in the difficult manoeuvres, a higher number of position checks was performed during the Approach_End and the Swing coding window (both significantly higher than the checks performed during the Approach_Start). These results further support the idea that when tolerances are reduced around the vessel (the difficult condition implied less room available for the vessel to swing), more position checks are required to be performed to ensure the success of the manoeuvre.

**Direction Check**

Direction Check is another important sequence of gaze behaviours that enables pilots to appreciate changes in ship direction, and consequently to keep the vessel in a suitable and safe position. Monitoring the relative motion of the bow relative to objects in the background helps pilots to identify the presence of any rotation (yaw), which influences the vessel's direction of motion. We found that the frequency of Direction Checks significantly decreased from Approach_Start, at the very beginning of the manoeuvre, to the Swing and the Closing. This finding is contrary to the trend observed for the variable Position Check, where checks increased as pilots progressed throughout the manoeuvre. At the beginning of the manoeuvres, the vessel was sailing at an appropriate speed. One of the first concerns of the pilots was to check the direction towards which the vessel was heading. An incorrect course would necessarily mean that the future position of the vessel could be incorrect. During Approach_Start, pilots seemed to focus their attention more around the bow and the equipment that indicated the heading of the vessel. During Approach_End and particularly during the Swing and the Closing, their priority changed, with pilots becoming more and more concerned about establishing their correct position by gazing in different directions and attending less to the bow. Pilots still looked in the direction of the bow, but this was mainly to determine if distances and clearances from obstructions were maintained. Evaluation of the Closing revealed a significant interaction between the factors Difficulty and Coding Window, where the number of checks increased in the easier manoeuvres and reduced in the harder ones. This may depend on the fact that in the easier manoeuvres the vessels (being rotated in a bigger basin) generally had to be driven alongside to the berth for a longer distance, requiring again focus on the bow.

**Speed Check**

To gain an appreciation of how fast a vessel is moving, pilots can simply observe the speed of objects located perpendicularly to the direction of vessel's movement. That is why this sequence included gazes at the beam, in combination with gazes at speed instruments. Having the capacity to correctly appreciate and maintain control of vessel's speed is crucial to successfully performing a safe rotation in confined waters. The general trend found was an increase in frequency of speed checks in all the manoeuvres, with Approach_Start having a significantly lower number of checks compared to the other coding windows. To moor safely a vessel alongside a berth, pilots have to progressively reduce the speed of the ship in order to arrive alongside with minimal momentum. The lateral monitoring, while using the log of a ship (which provides information about the longitudinal and lateral speeds), allows the pilot to calibrate the landing until the final touch on the fenders which is particularly important at the end of the manoeuvres. Figure 5 shows also an interaction between the factors Difficulty and Coding Window. It can be noticed how in the difficult manoeuvres a higher number of checks was performed during the Approach_End compared to the easy one.
There was actually a significant main effect for the factor Difficulty, with difficult manoeuvres recording higher shores than the easy ones. The effect size though was very small, drawing our attention more on the interaction in the Approach_End, than the difference of the whole manoeuvres between the two conditions.

## Conclusions

This study explored the opportunity to consider not simply single objects or areas of interests as targeted by gaze, but entire gaze sequences. Gaze sequences, as clusters of organized overt behaviour, can be related to underlying complex tasks and goals. Pilots adopted different gaze patterns to respond to the specific challenges created by each manoeuvre (hard and easy) during each phase of the berthing. The findings showed that pilots were engaged in goal-directed action (i.e., they had different shiphandling priorities during different phases of the manoeuvre). Having internal detailed knowledge (Schemata) about vessel's response and manoeuvrability in different conditions (Object), the pilot prioritised attention on different sources of information, when more relevant (Exploration). For example, in the approach, the focus on speed was more relaxed since associated with broader maximum / minimum speed limits necessary to counteract the environmental conditions, while during the closing the speeds were monitored far more closely, to land correctly on the fenders without damage.

Measuring overt behaviour such as gaze patterns, can be used as a relatively unobtrusive tool to measure attention, and as such inform our understanding of internal schemata. Gaze behaviour can then be related to performance outcomes. The use of gaze measurements could offer the opportunity to verify which of those patterns result in the best outcomes, guiding future improvements as better described in the next paragraph.

### Future Applications and Added Value.

Examination of the most relevant and useful source of information used by pilots when undertaking manoeuvres, depending on context and conditions, can assist designers and manufacturers to optimise equipment designs, and trainers to teach more efficient and appropriate shiphandling techniques. In previous research we provided an example of how it is possible to obtain shiphandling performance outcomes (Luca Orlandi et al., 2015). Being able to define and monitor meaningful gaze behaviours through the use of eye trackers, could be used to improve evaluation of training outcomes, actual performance and, in the future, real time activities in normal working environments. Future studies involving gaze analysis (and therefore attention) specifically with reference to the source of information (electronic equipment, external visual aids..) preferred by pilots (Itoh et al., 1990) could offer important insights regarding information resource management and shedding preferences once task demand begins to overcome pilot capabilities (M. S. Young et al., 2015). All of the above will be useful in the reconstruction of accidents within simulated environments – an increasing practice in the maritime domain.

As this research was conducted in a simulated environment, future studies should also consider the collection and comparison of similar data in a real working environment. For this to be achieved, it will be critical for data collection to be robust and unobtrusive, in order not to distract or interfere with berthing operations. Should that data collection be possible, it could provide a better understanding of normal and abnormal, individual and group behaviours, which could help to identify critical operations and levels of performance. Once these behavioural patterns were identified, they could be exploited as prodromal indicators of critical conditions or good performance. Beyond this, such data collection will provide a better appreciation of the realism of the simulated environment through a comparative analysis of the same behavioural patterns.

Importantly, this study informs improvements in Australian national pilotage standards (ATC, 2008) around issues such as the use of simulation facilities for training and continuing professional development of pilots. These include use of the experimental protocols to update training materials and standards, around the relationship between behavioural patterns and performance outcomes, with the aim of enhancing the safety of shiphandling techniques.

**Acknowledgements**

We would like to thank the Australasian Marine Pilots Institute and Smartship Simulator for the priceless support offered in the realization of this work. We are deeply grateful to the Pilots that patiently bore with the intervention of too many obnoxious devices!

# Appendix A

**Missing Data.**

| | Homeport | | | | | | | | Foreign | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Easy | | | | Difficult | | | | Easy | | | | Difficult | | | |
| Subject | Appr_Start | AppEnd | Swing | Closing | AppStart | AppEnd | Swing | Closing | AppStart | AppEnd | Swing | Closing | AppStart | AppEnd | Swing | Closing |
| 1 | | | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | | M |
| 4 | | | | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | | | | M |
| 6 | | | | | | | | | | | | | | | M | M |
| 7 | | | | | | | | | | | | | | | | |
| 8 | | | | | | | M | M | | | | | | | | |
| 9 | | | | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | | | | | |

The letter ‘M’ in the table identifies those cases were data was missing due to clashes.

## 6.5. Paper IV – Unpublished Version

# Interpreting changes in marine pilots' perceptual cycle through gaze detection and speech patterns.

Luca Orlandi (1), Benjamin Brooks (1), Joanne Wood (2) and Alexander Black (2)

(3) *University of Tasmania, Australian Maritime College, National Centre for Ports and Shipping, Launceston, Tasmania 7250, Australia.*
(4) *Queensland University of Technology, Faculty of Health, School – Optometry and Vision Science.*

(E-mail: luca.orlandi@utas.edu.au; Tel: +61 3 6324 3999)

## Abstract

This paper explores the perceptions of shiphandling by analysing visual and speech behaviour of a cohort of marine pilots engaged in berthing manoeuvres. Four moorings, which varied in levels of difficulty, familiarity with the port and phase of the manoeuvres, were completed in a full mission bridge simulator. Video clips obtained with the use of eye tracking goggles worn by pilots were screened and coded. Pilots' gaze and speech behaviour was labelled, adopting specific behavioural markers, and analysed as combinations of meaningful sequences. These markers were scored on four coding windows of 5 minutes each, purposely placed into each one of the phases of the four manoeuvres. The adopted sequences of behavioural markers showed significant results with respect to the different manoeuvring phases, levels of difficulty and familiarity with the port, that required pilots to adopt different shiphandling techniques and highlighted significant changes in the frequency of checks of vessel's position, direction and rotation. Significant differences were also recorded in the number of orders and gazes at a passage plan. The development of this methodology will enable further exploration of marine pilot expertise and demonstrates how objective and unobtrusive measures can be defined and adopted to quantify shiphandling capabilities, improving training and reducing accidents.

## Practitioner Summary

Gaze and speech behaviour were investigated in a cohort of marine pilots during mooring simulations. Combinations of defined behavioural markers showed significant differences in the execution of different shiphandling manoeuvres. Application of this method with a larger sample may provide objective and unobtrusive measurements, which can define and quantify situational awareness and expertise in this occupational group. Implications for training and operations are identified.

## Keywords

1. Shiphandling performance 2. Eye trackers 3. Ship Simulator 4. Marine Pilotage 5. Behavioural Markers.

## Introduction

Shipping is the predominant transport mode for the movement of freight in the world. Commercial vessels carry around 90% of the world trade. The maritime industry generates an annual income of over half a trillion US dollars in freight rates, with a worldwide population of 1,187,000 seafarers serving on internationally trading merchant ships (ICS, 2016). In contrast to this, seafarers and ships generally remain out of sight and out of mind from both the general public and the research community. Few people interact with ships (compared to cars, trains and aeroplanes) and their levels of safety only breach our awareness during major accidents such as the Costa Concordia grounding near the island of Giglio, off the Italian coast.

This ‘hidden’ mode of transport combines with a highly litigious environment (governments will often imprison seafarers after incidents), to further obscure our understanding of human performance in this domain. However, the nature of ship-handling should be of interest to human factors researchers. A ship can be more than 300 meters long, weighing more than 250,000 tonnes, operating in dynamic moving oceans with complex hydrodynamic effects. The complexity of the interaction and sheer size indicates an environment ripe for the study of perception and the broader concept of situational awareness.

### *Situational Awareness Defined*

Among the many definitions of Situational Awareness available in the literature, probably the most commonly adopted is: “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future” (M. R. Endsley, 1988). From this definition, Endsley’s three level information processing model of situational awareness was constructed. This model explains how projection to the future can only begin from the understanding of an actual state. Such situation awareness is based on the constant flow of perceptions that provide the subject with elements to build such a construction. If perceptions are partially or wrongly interpreted, they will lead to inaccurate understanding of the actual state, implying incorrect projection into the future. This consideration is particularly relevant for marine pilots. Although ships are moving slowly in comparison to other forms of transport, the size and hydro-dynamic properties of the ship require the pilot to be able to project forwards in time and space, given that a bulk-carrier of a displacement of 110000 tonnes takes more than 4 km to come to a complete stop. For this reason it is crucial for pilots to build and maintain an accurate understanding of the actual situation, knowing that, to achieve this, it becomes even more critical to receive and interpret the most relevant and accurate information (perceptions) on an ongoing basis. This study focuses specifically on the evaluation and quantification of this fundamental process of information acquisition. In addition, appreciating the cyclical nature of such process, this study favours a different model the Smith and Hancock (1995) Perceptual Cycle Model of Situation Awareness, which is considered more relevant to this situation than that of Endsley.

### *Perceptual Cycle Model of SA*

This study was based on the use of the Perceptual Cycle model of SA developed by Smith and Hancock (1995). The model views SA as “a generative process of knowledge creation and informed action” (Salmon et al., 2008). The implied process of information gathering is cyclical and the beginning and end of the process are continual. The model is based upon Niesser’s Perceptual Cycle (1976), which describes how individuals interact with the world and how they use schemata to modify their interactions with the world. This dynamic approach to achieving SA (Uhlarik & Comerford, 2002) consists of three elements:

- The object (i.e. available information in the external environment).
- The schemata (i.e. the internally structured knowledge that is developed through training and experience, and is stored in long term memory when not in use).
- Exploration (i.e. a search of the environment by the observer).

Internally held schemata (knowledge) direct a person's exploration of the world. These mental models facilitate anticipation of situational events directing the person's attention to cues in the environment and their eventual course of action. The person then carries out checks to confirm that the evolving situation conforms to their expectations (K. Smith & Hancock, 1995). The outcome of such exploration may modify the original schemata, which in turn directs further exploration. This process continues in a cyclical manner, obtaining SA as the dynamic result of the interaction of the person with the world; any unexpected events prompt further search and exploration and in turn modify the individuals existing model (N. Stanton et al., 2013).

***Using Eye-Tracking to Inform our Understanding of Situational Awareness***

Data collected in eye-tracking studies can be both interpreted by, and used to inform, our understanding of concepts, such as situational awareness. In particular, measurement of gaze position within a natural environment through the use of eye-tracking provides important information regarding where overt visual attention is directed.

Eye trackers have been extensively utilised for assessing and training in several fields, as well summarised in recent reviews on the topic (Gegenfurtner et al., 2011; Hermens et al., 2013; Rosch & Vogel-Walcutt, 2013; Tien et al., 2014). In the transportation industry, several examples can be found where eye trackers have been employed to analyse gaze behaviour in different contexts and conditions. Studies conducted in the automobile industry have investigated, for example, differences between experts and novices (Di Stasi et al., 2011; Falkmer & Gregersen, 2005; Fisher et al., 2007), glance differentiation and distribution while executing in-vehicle tasks (Victor et al., 2005) and drivers experiencing specific medical conditions or visual impairments (Lee et al., 2016). A large body of work has also been conducted in the airline industry, studying the gaze behaviour of pilots in cockpits (Sarter et al., 2007; Weibel et al., 2012).

As indicated by earlier comments regarding the 'hidden' nature of maritime transport, only a few gaze behaviour studies have been conducted in the maritime industry. One study compared the gaze behaviour in 16 experienced and novice fast boat drivers and their use of navigational aids (Forsman et al., 2012). This study identified that novice drivers looked more at objects that are close to themselves, like instrumentation within the cockpit, while the experienced look more at objects further away from the boat. Further, novice boat drivers used the electronic navigational aids to a larger extent than the experienced, especially during high speed conditions. Research has also been conducted in simulated maritime environments, where the use of eye trackers has been adopted to compare novice and experts engaged in overtaking and bypassing a ship in a narrow canal (Muczyński et al., 2013). This study also demonstrated that experienced captains directed their gaze differently compared to novice shiphandlers. Experienced captains focused mainly on visual observation of the relative positions of the involved ships and assessment of the distance between them, and spent less time monitoring own ship conning equipment.

An early study investigated differences in gaze distribution between student and experienced navigators in a simulator using 6 different participants, 2 experienced navigators and 4 students (M. Lützhöft & Dukic, 2007). Results showed that experienced navigators made less glances per minute and generally followed a well organised scanning pattern.

To date, limited research has been carried out specifically on Marine Pilots using eye trackers. The visual perception of marine pilots is, however, of particular interest because the size of the ships that they manoeuvre requires particularly acute spatial awareness. A qualitative study evaluated video recordings from simulated manoeuvres performed by 12 pilots conducting ships in and out the port of Oslo (Hontvedt, 2015). The ships used for the simulations were “azipod” vessels, meaning that they were provided with two propellers on the stern to rotate the direction of their thrust on 360 degrees. Due to the specific characteristics of this propulsion system, pilots required dedicated training in ship simulators to better understand their use and implications. Five sections in the manoeuvres were selected and analysed to provide different examples of interaction between pilots and ships’ crew while conducting the vessel. The study did not provide any quantitative outcomes, but offered examples of interaction analysis: an empirical and video-based method was used to study social interaction, as it evolves through talk, non-verbal interactions and the use of artefacts and technologies.

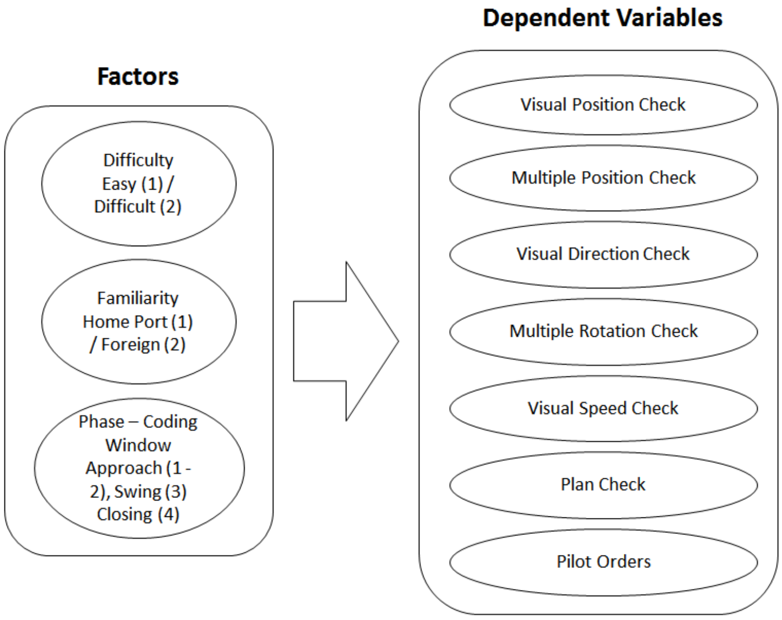
A more recent study compared eye tracking data collected on board a littoral combat ship and in a bridge navigation simulator (Hareide et al., 2016) and reported that gaze behaviour was similar in the two environments with respect to gaze dwelling time on specific areas of interests. This highlights the potential value of assessing pilots in simulator environments and that eye tracking data can help to identify limitations and drive improvements in human machine interfaces.

***Research Context***

This study investigated how different manoeuvring conditions influenced gaze and speech behaviour of a group of marine pilots while berthing ships in a simulator. Marine Pilots are ship’s captains that are specifically trained and certified to manoeuvre vessels within critical coastal and port waters. They embark a ship outside port waters and then work with the bridge team to navigate the ship to berth. While ship’s Captains still retain full charge or responsibility of the safety of the vessel, pilots generally take the “conduct”, manoeuvring the ship until a safer position is reached or the vessel is alongside the assigned mooring.

Ships’ pilotage implies complex interactions with a bridge team, tug masters, a vessel traffic service and electronic equipment. This study evaluated and compared measurements of specific gaze and speech patterns (orders) while pilots were involved in different shiphandling manoeuvres, with different levels of difficulty, in different port environments and during different phases. A set of Behavioural Markers (BM) were defined in order to code video clips obtained from eye trackers worn by participants and to create a list of meaningful dependent variables that could be used for statistical analysis. The study hypothesis was that different types of manoeuvres and different stages of these manoeuvres would require different behavioural responses. Studying and understanding the link between those responses (in terms of occurrence of specific BM) and shiphandling outcomes, could help to better evaluate pilots’ expertise (Luca Orlandi et al., 2014) for safety, training and assessment purposes. The list of independent and dependent variables is represented diagrammatically in Figure 1.

**Figure 1. Independent and Dependent Variables.**




## Methodology

### Participants

Participants included 10 marine pilots, employed by an Australian pilot company. In total, 40 individual pilotage 'runs' were recorded producing 800 minutes of video available for coding. The marine pilots were all males in good health, as required by national professional medical standards set by the Australian Maritime Safety Authority, which involved a biannual (or yearly, if over 55) full medical check (AMSA, 2010). An Analysis of Variance (ANOVA) for age and service confirmed no significant difference between the participants and the rest of the pilot population working for the same company. All the pilots involved in the research had more than ten years of previous experience as a qualified pilot. The research complied with the ethical standards of our University and the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

### Experimental Design Factors

To investigate pilots' perceptual cycle, different berthing manoeuvres were set as experimental conditions. Each manoeuvre included the whole process necessary to transfer the ship from a defined initial position to a berth within constrained port waters, with the use of own and/or external means of propulsion (i.e. tug boats to assist, when allowed). The manoeuvres were presented to pilots before being performed in the simulator, since every participant was required to provide a plan, such as would be normally discussed by pilots and ship masters before a ship enters into a port (Wild & Constable, 2013). In each of those manoeuvres, three main factors were manipulated.

**Table 1. Experimental Design – List of factors per manoeuvre.**

| | ***Manoeuvre*** | | |
|---|---|---|---|
| | ***(b) Level of Difficulty*** – Easy "1" or Difficult "2" | | |
| ***(a) Port Familiarity*** - Home "1" or Foreign port "2" | ***(c) Phase*** Approach "1" | ***(c) Phase*** Swing "2" | ***(c) Phase*** Closing "3" |

These three main factors were: (a) "port familiarity" (from now on referred as "port"), (b) "difficulty", and (c) "phase". The first factor, "port", took into account whether the manoeuvres were conducted in the participant pilot's homeport (the port where they were regularly working) or in a foreign port. The foreign port was a virtual port only present in the simulator software. This port was chosen to avoid any possibility of a learning effect associated with previous manoeuvring experience the participants may have possessed and to provide support for methodology reliability. The pilots' homeport is coded "1", while the virtual port is coded "2", in the tables presented in Appendix B. The second factor was the ship-handling level of "difficulty". To control the level of difficulty, the parameters of specific manoeuvres were altered as summarised in Table 2. The easy level is coded "1" while the difficult level is coded "2".

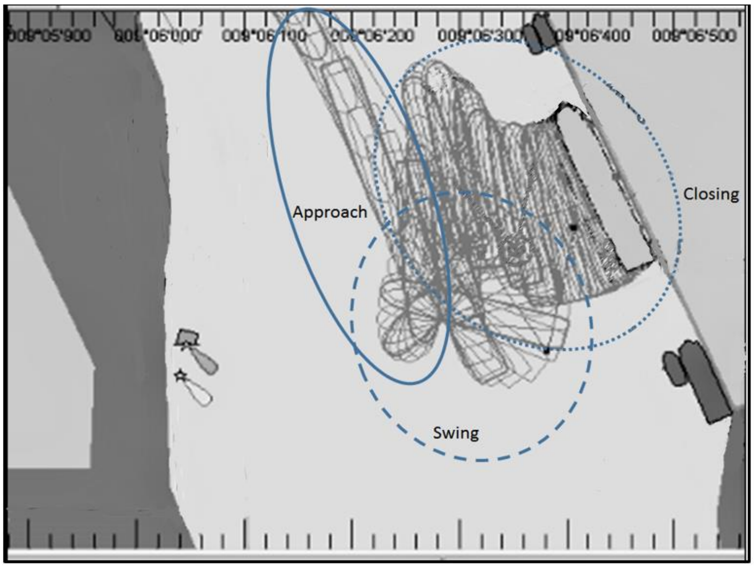
**Table 2. Levels of Difficulty – Adopted in both Ports.**

| | **Pier - Spatial constraints** | **Environmental conditions and forces** | **Vessel characteristics** | **Tugs** | **Interactions with traffic** | **VTS Comms (1)** |
|---|---|---|---|---|---|---|
| **Level 0** **Easy** | Big Swing Basin (3 times Vessel LOA (4)) | Current: 0.7 Knt Wind: 15 Knt Good Visibility | Single Controllable Pitch Propeller (2) Bow Thruster (3) | None | 1 Interacting but not Interfering vessel | General Info No reporting Points |
| **Level 1** **Difficult** | Small Swing Basin (5) (1,5 times Vessel LOA) | Current: 2 Knt Wind: 25 Knt Poor to no Visibility - Heavy Rain | Single Fixed Pitch Propeller No Thrusters | As required by Pilot | 1 Interacting 1 Interfering vessels | General Info and Traffic Advice Reporting Points |
| **Notes** | (1) Vessel Traffic Management station present in a port and managing ships via radio communications; (2) Propeller capable to change the water thrust direction changing the angle of the blades instead of direction of rotation; (3) A thruster is a propeller positioned perpendicular to the ship keel axis. Placed on the bow or on the stern, induces transversal / angular motion; (4) Length Over All, maximum length of a vessel; (5) Wider area, within constrained waters, where ships have sufficient room to rotate and revert their direction. | | | | | |

Level 1 provided a comparable level of difficulty to that of routine operations. Level 2 aimed to engage pilots with a scenario slightly exceeding the safety limits established in the pilots' homeport, without losing construct validity.

Each manoeuvre required pilots to complete a mooring using the side of the ship opposite to the berth position on commencement of the exercise. This implied that for each manoeuvre the ship had to swing (rotate 180°) before she could be moored. Each manoeuvre therefore had to be developed through three main phases. The three phases were: the “approach” (from the initial position until the start of the swing), the “swing” (from the start of the swing until the rotation was completed and stabilised), and the “closing” (from the end of the swing until a defined distance from the berth). Figure 2 is a screenshot taken from the simulator interface showing one of the four manoeuvres (the easy manoeuvre in the virtual port).

**Figure 2. Example of a manoeuvre as shown by the simulator interface. Three phases highlighted by circles.**




In summary, four different manoeuvres were conducted by each pilot. Two of the four manoeuvres were conducted in the pilots’ homeport (one manoeuvre for each level of difficulty), two were conducted in the foreign port (one manoeuvre for each level of difficulty). The same four manoeuvres were conducted by each participant, in random order, to minimise a possible learning effect. Spatial constraints due to port dimensions were purposely maintained to be as similar as possible, modifying the virtual port in order to match homeport dimensions as summarised in Table 3.

**Table 3. Proportions between vessels and port dimensions.**

| Ship | LOA (m) | Ratio between Ships | Breadth (m) | Disp (ton) |
|---|---|---|---|---|
| **Torm Laura (diff Lvl 0)** | 183 | 0.7 | 32 | 54925 |
| **Arcturus (diff Lvl 1)** | 269 | 1.45 | 48 | 143200 |
| **Ratio** | **Torm LOA** | **Torm Breadth** | **Arcturus LOA** | **Arcturus Breadth** |
| **Basin diameter (470 m)** | 2.6 | 14.7 | 1.7 | 9.8 |
| **Channel width (300 m)** | 1.6 | 9.3 | 1.1 | 6.2 |

The Maritime Safety Queensland Simulator located in Brisbane was used (Smartship® Simulator www.smartshipaustralia.com.au). This "Full Mission Bridge" simulator is classified as Class A (NAV) according to the standards issued by the classification society Det Norske Veritas Germanischer Lloyd (2014). It is capable of simulating a total shipboard bridge operation, including the capability for advanced manoeuvring in restricted waterways. Before the experimental manoeuvres, pilots were required to perform a very simple mooring with a vessel different from those used in the experimental runs. This first manoeuvre was used as a familiarisation run.

### Behavioural Markers

During those manoeuvres, eye movements and audio were recorded using a pair of light-weight eye tracking goggles (ASL® Mobile Eye XG® - www.asleyetracking.com). This eye-tracker system comprises two cameras each sampling at 30 Hz: a forward facing scene camera and an eye camera, capturing the infrared corneal reflection and pupil position of the right eye. A calibration procedure, which determines where gaze is located within the scene, was performed at the beginning of each manoeuvre. The eye-tracking video and audio recordings obtained were coded (Kipp, 2010) using a set of labels, that identified specific elements that the shiphandler directed gaze towards or verbally reported in the simulated environment. A list of all the analysed sequences obtained that combined different behavioural markers is reported in Appendix A.

Due to the length of time necessary to complete the manual coding of the whole duration of the manoeuvres, a sampling strategy was applied. A total duration of 20 minutes of video coding was obtained for each manoeuvre, considering four different sections of five continuous minutes. The coding windows' location in each manoeuvre was chosen based on specific criteria. Each coded section had always a minimum duration of five minutes (unless the whole duration of that specific manoeuvring phase was less than this). In the swing and the closing phases, the coding window was always placed in the middle. In the approach phase, the coding windows were placed one at the very beginning and the other one at the end, finishing with the beginning of the swing section. Time gaps between coded sections, varied according to the duration of the different phases. Performance during these chosen coding windows was analysed based on the occurrence of the previously described sequences of behavioural markers (BM). Those sequences were derived either from shiphandlers gaze distribution on different objects or from combinations of BM such as gazes and communications (i.e. the communication of a rudder order and then the visual check of the rudder indicator).

**Table 4. Locations of video coded sections in each manoeuvre**

| Approach | | | Swing | | | Closing | | |
|---|---|---|---|---|---|---|---|---|
| Coding Window 1 | GAP | Coding Window 2 | GAP | Coding Window 3 | GAP | GAP | Coding Window 4 | GAP |

Scores in all of the coding windows from each participant and from each manoeuvre (see table 4) were processed. The maximum frequency recorded in each coding window was chosen for each of the identified sequences (dependent variables). A data table of a total of 160 cases (10 subjects X 4 manoeuvres X 4 coding windows) for each dependant variable was obtained and analysed. Coding windows where chosen specifically during these phases, given that they are extremely different from a shiphandling point of view. Our expectation was that there would be significant differences between these coding windows and our aim was to better describe and characterize such differences.

### Statistical Methods

In order to determine if the chosen sequences of BM were able to depict different patterns during the execution of the different manoeuvres, a 3 way repeated measures ANOVA (by subject) was performed using the statistical package IBM SPSS (IBM_Corp., 2010). Main effects and interactions were included. For significant 3-way interactions, the use of tests of simple main-effects were explored using 2-way and 1-way ANOVAs. The factors were "difficulty", "port" and "coding window" (see Table 4). The assumption of normality of distribution and heteroscedasticity of the analysed scores were tested, as well as the presence of outliers (visually inspecting boxplots). Significant results obtained on the dependant variables are reported and discussed in the following sections:

## Results and Discussion

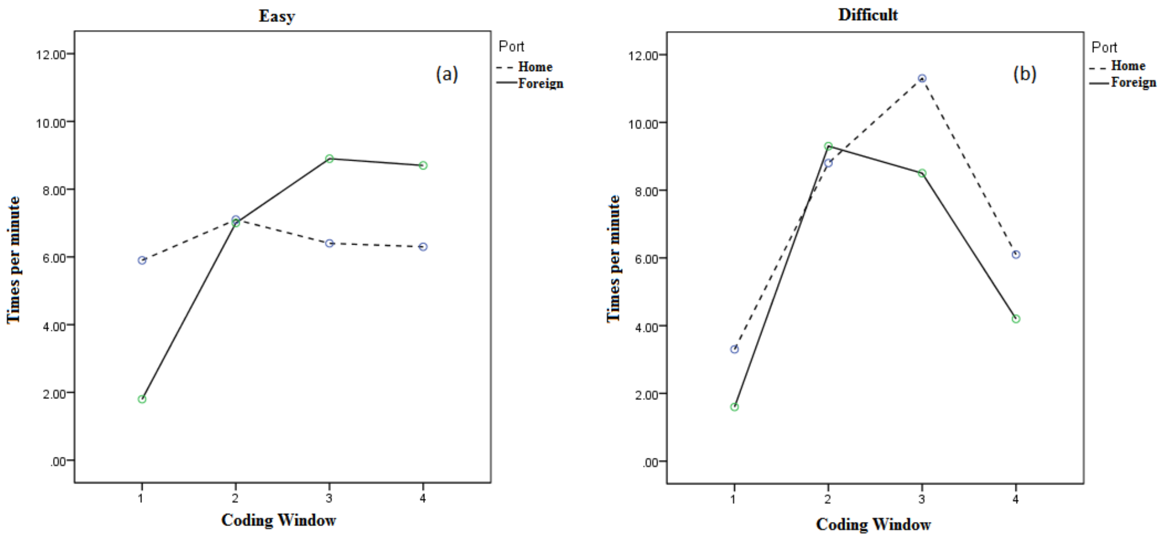
### Missing Data

During the execution of the 40 manoeuvres, a total of four crashes were recorded. A "crash" was defined as an impact or a grounding that required the interruption of the simulation and data collection. All of the crashes were experienced during the more difficult manoeuvres (Diff = 2), three during the manoeuvres in the foreign port (Port = 2) and one in the homeport (Port = 1). Given that the simulator had to be stopped and reset after a crash, no further data collection for that manoeuvre was possible. More specifically, two crashes happened during the approach (first phase of the manoeuvre), so no data could be collected during the following two phases (swing and closing). Two crashes were recorded during the swing, so no data could be collected for the following phase (closing). Three impacts were also experienced (one in the foreign port in easy manoeuvres and two during the swing in both ports in the difficult manoeuvres). An "impact" was classified as a contact of the vessel with another ship or port infrastructures that did not impede the continuation of the manoeuvre. The unavailable data was left missing in the dataset used in the following analysis.

### Visual Position Check

This behavioural marker reports the number of times per minute the pilots gaze was directed towards an external object and then moved to another object, that was at approximately 90 degrees from the bearing of the previous one. Sequentially comparing objects at approximately 90 degrees from each other, is a technique that pilots use to gain their understanding of where they are located in the manoeuvring space. As reported in Table B.1. in Appendix B, significant results were found both as main effects and as interactions between factors.

**Figure 3. Graphs of Visual Position Check**

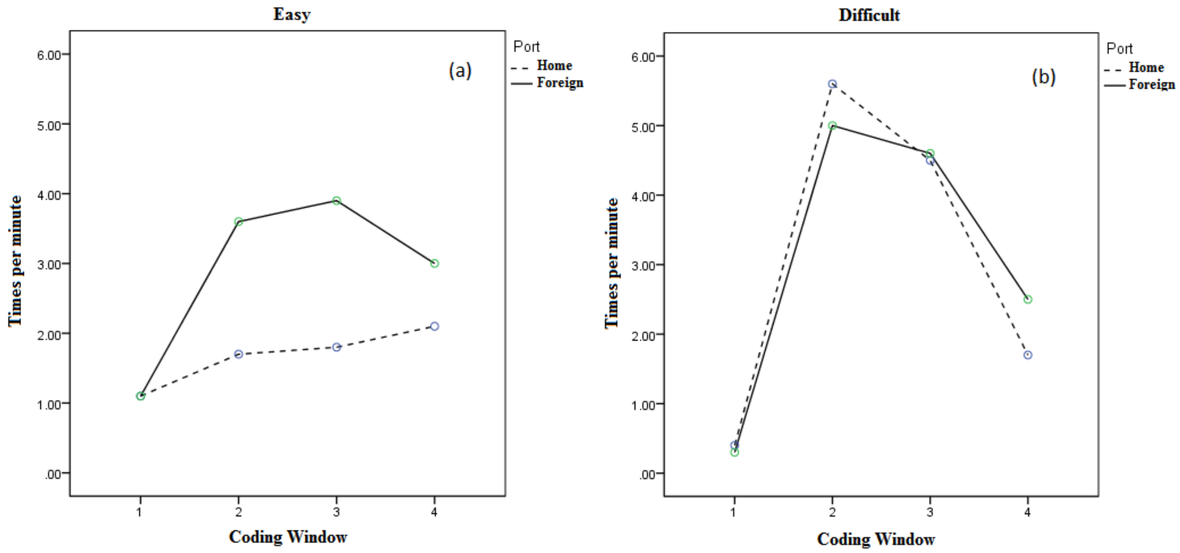



As shown in Figure 3(a) and 3(b), the frequency of this behavioural measure significantly increased from Approach_1 (CdWd_1 = 3.15), at the very beginning of the manoeuvre, to Approach_2 (CdWd_2 = 8.05) and Swing (CdWd_3 = 8.78). This significant 3 way ANOVA main effect, was confirmed by separate 1 way ANOVA (fixing factors Port and Difficulty). The only exception was for the easy manoeuvre in the homeport (figure 3(a) dotted line), where there is no significant effect on the factor coding window. This trend can be explained by the fact that, at the beginning, when the exercise is started, the ship is in a known position and making way, so for pilots it becomes more important to monitor the direction where the ship is heading instead of checking its actual position. Later on in the manoeuvre (even though always important), it becomes particularly relevant for the pilot to be confident about the exact position of the ship. Coding window 2 is when and where the pilot decides to start the swing. He has to be sure that the ship is in a suitable position that would allow enough clearance to rotate with enough distance from surrounding obstructions. The frequency of Visual Position Check reaches its peak during the swing (coding window 3). In addition, as can be observed by comparing figure 3(a) with 3(b), the peak is significantly higher for the difficult manoeuvres, (where the space allowed is much less). Notice also how in figure 3(b), lower values are recorded for the Closing (CdWd = 4, which may depend on the 4 crashes that all occurred during the difficult manoeuvres, which precluded data collection in those and the following phases.

### Multiple Position Check

This behavioural marker refers to when the pilot, in addition to the previously described Visual Position Check, also check the electronic equipment which might include, for example, the radar screen or the electronic chart plotter. In Table B.2. in Appendix B, only significant results are reported. Figures 4(a) and 4(b) illustrate the significant interactions highlighted by the results reported in Table C.2.:

**Figure 4. Graphs of Multiple Position Check**

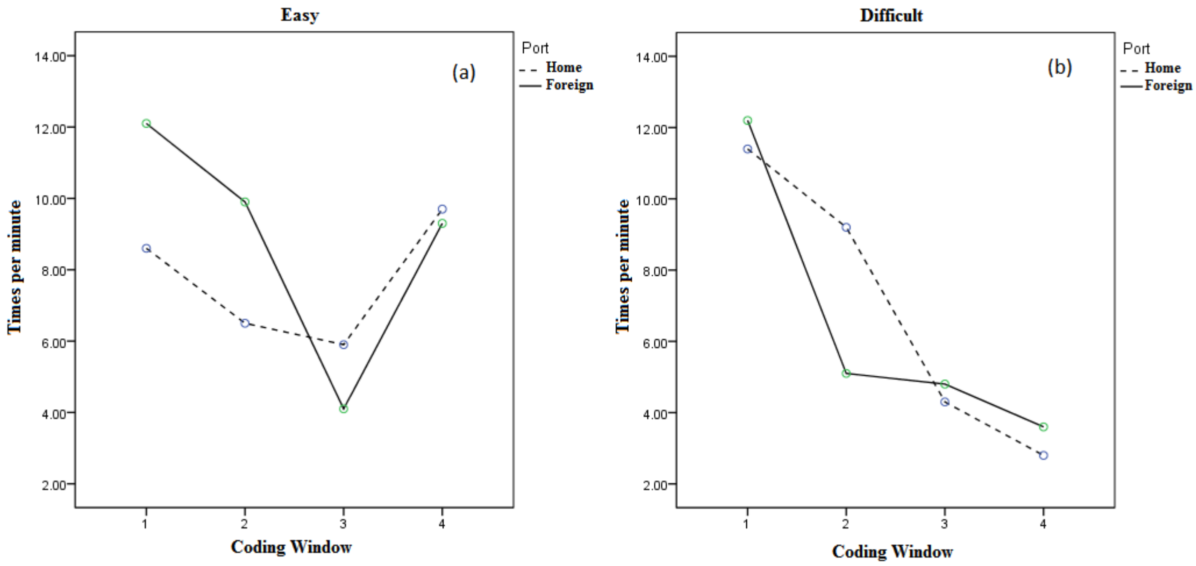



As previously highlighted with the behavioural marker Visual Position Check, the Multiple Position Check also showed a generalised and significant increasing trend from Approach_1 until the Swing (CdWd_1 = 0.73, CdWd_2 = 3.98, CdWd_3 = 3.70), with the exception of the easy manoeuvre in the homeport. A drop in the Closing coding window (CdWd_4 = 2.33), is significant for both difficult manoeuvres (see figure 4(b)). Figure 4(a) clearly shows how in the easy manoeuvres (Diff = 1) there was significantly less use of the electronic positioning equipment in combination with the visual checking in the homeport (dotted line) compared to the other manoeuvre (continuous line) (Port_1 = 1.68, Port_2 = 2.90). The homeport also had statistically lower scores for the difficulty factor (2 ways ANOVA, fixed factor Port = 1) compared to the foreign port (Diff_1 = 1.68, Diff_2 = 3.05). This suggested that familiarity with the homeport could have enabled the pilots to rely more on their capability to refer to known features observed in the environment, than on the positioning equipment.

### Visual Direction Check

The behavioural marker Visual Direction Check recorded the number of times per minute that the pilots' gaze was moved between an external object within 30 degrees centred on the bow and another object in the same sector and / or a heading instrument (gyro repeater). This represents a strategy through which pilots may perceive and monitor vessel direction of motion. As reported in Table B.3. in Appendix B, significant results were found both as main effects and as interactions between factors.

**Figure 5. Graphs of Visual Direction Check**

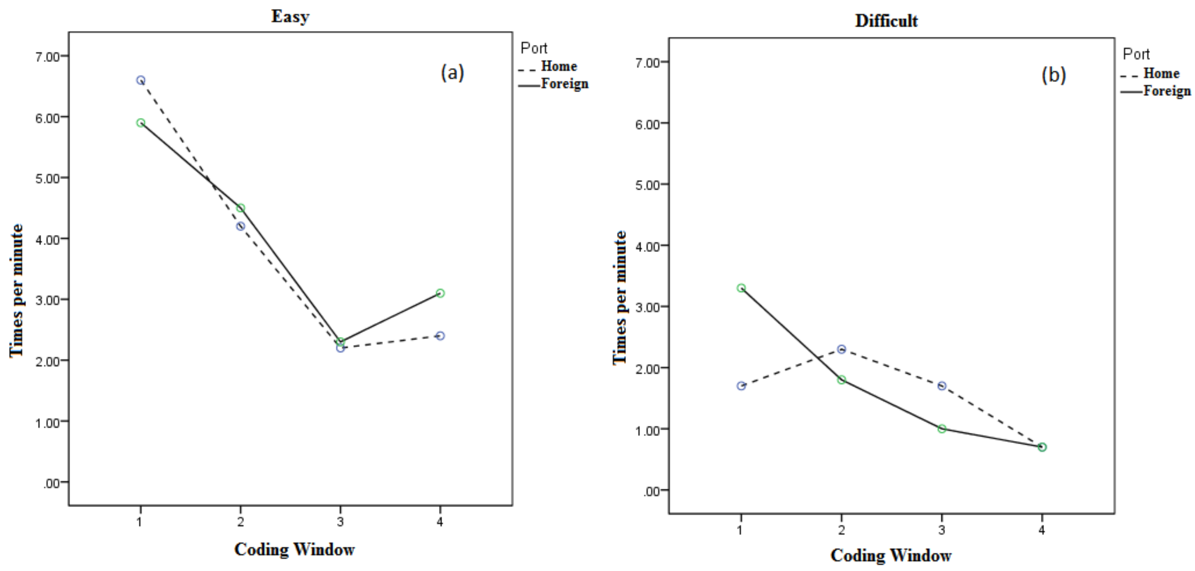



In figure 5(a) and 5(b) frequency (times per minute) significantly decreased from Approach_1 (CdWd_1 = 11.08) and Approach_2 (CdWd_2 = 7.68) (at the very beginning of the manoeuvre) to the Swing (CdWd_3 = 4.78). This finding is directly opposite to the trend observed for the BM Visual Position Check. At the beginning of the exercises the vessel is sailing at an appropriate speed, implying that one of the first concerns of the pilots is to check that the direction the vessel is heading is correct. At the beginning, the priority is to monitor and correct if the direction of advance is incorrect, since an incorrect course will necessarily mean that the future position of the vessel will be incorrect. During Approach_1, pilots seemed to focus their attention more around the bow and the equipment that indicated the heading of the vessel. During Approach_2 (coding window 2) and statistically more during the Swing (coding window 3) their priority changes. Pilots become more and more concerned about establishing their correct position by gazing in different directions and attending less to the bow. Pilots still look in the direction of the bow, but this is mainly to appreciate if distances and clearances from obstructions are maintained. Evaluation of the Closing reveals a significant interaction between the factors Difficulty and Coding Window, where the scores increase in the easier manoeuvres and reduce in the more difficult ones. Again, this may depend on the crashes occurred and the data lost or also on the fact that in the easier manoeuvres the vessels (being rotated in a bigger basin) generally had to be driven alongside to the berth for a longer distance, requiring again focus on the bow.

## Multiple Rotation Check

The behavioural marker Multiple Rotation Check recorded the number of times per minute that pilots monitored the rate of turn of the vessel (ROT is the vessel yawing). It was considered a complete sequence when the gaze was shifted alternatively from an external object maximum 30 degrees off the bow and then directed back on the bow, combined with a visual check on the ROT sensor. This behavioural pattern may allow pilots to detect and correct any unwanted rotation of the vessel. As reported in table B.4. in Appendix B, in the 3 way ANOVA only main effects were found as significant outcomes on the factors of difficulty and coding window.

**Figure 6. Graphs of Multiple Rotation Check**

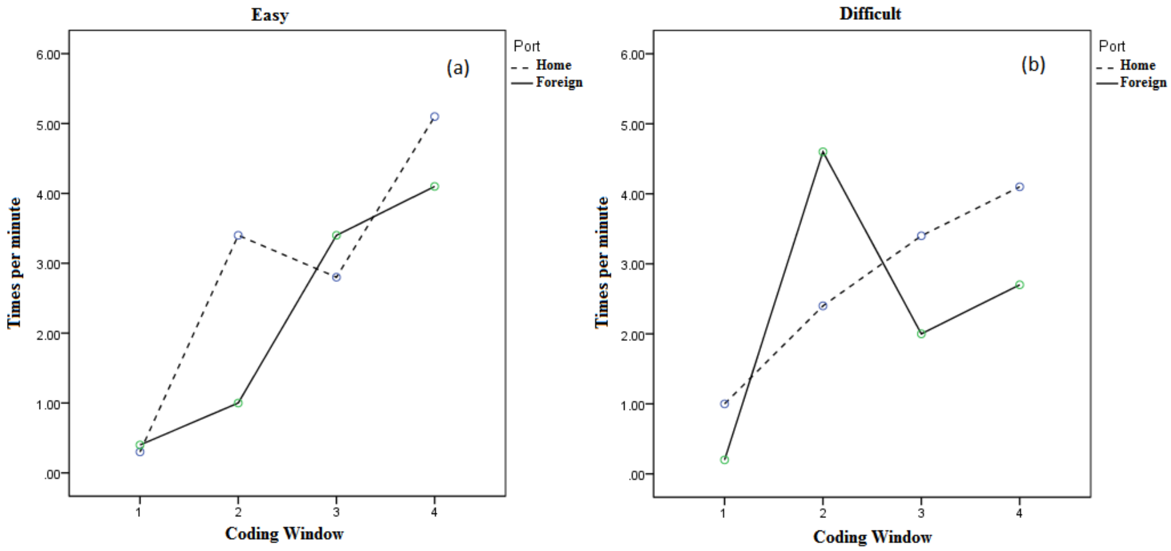



The main effect of the factor Difficulty is evident by comparing figure 6(a) and figure 6(b). It is evident how in the easier manoeuvres more checks were performed by pilots with reference to the rate of turn of the vessel (Diff_1 = 3.90, Diff_2 = 1.65). It should be noted that the vessel employed in the easy manoeuvres had a controllable pitch propeller (CPP). This propeller changes the angle of its blades to change the power of water thrust or even revert it. Pilots had to reduce the propeller blade pitch to a 0 angle, in order to reduce the ship speed, at the same time reducing the effectiveness of the rudder shielded by the rotating propeller. In doing so, pilots had to closely monitor the effect that this reduction of speed would have had on the ROT of the vessel, in order to correct it accordingly. The other main effect was recorded on the factor Coding Window. Figures 6(a) and 6(b) clearly shows how the checks on the ROT decreased throughout the whole duration of the manoeuvres, having their peak at the beginning (CdWd_1 = 4.38, CdWd_3 = 1.80); the slight increase in the easy manoeuvres in the Closing is not significant.

### Visual Speed Check

The behavioural marker Visual Speed Check recorded the number of times per minute that the pilots gaze was alternatively fixed on an external object at the beam of the vessel and then directed on a speed sensor (Log). Pilots may be able to determine the speed at which the vessel is moving, by monitoring the relative motion of objects at approximately 90 degrees to vessel motion, while reading a speed indicator. As reported in table B.5. in Appendix B, significant results were found both as main effects and as interactions between factors.

**Figure 7. Graphs of Visual Speed Check**

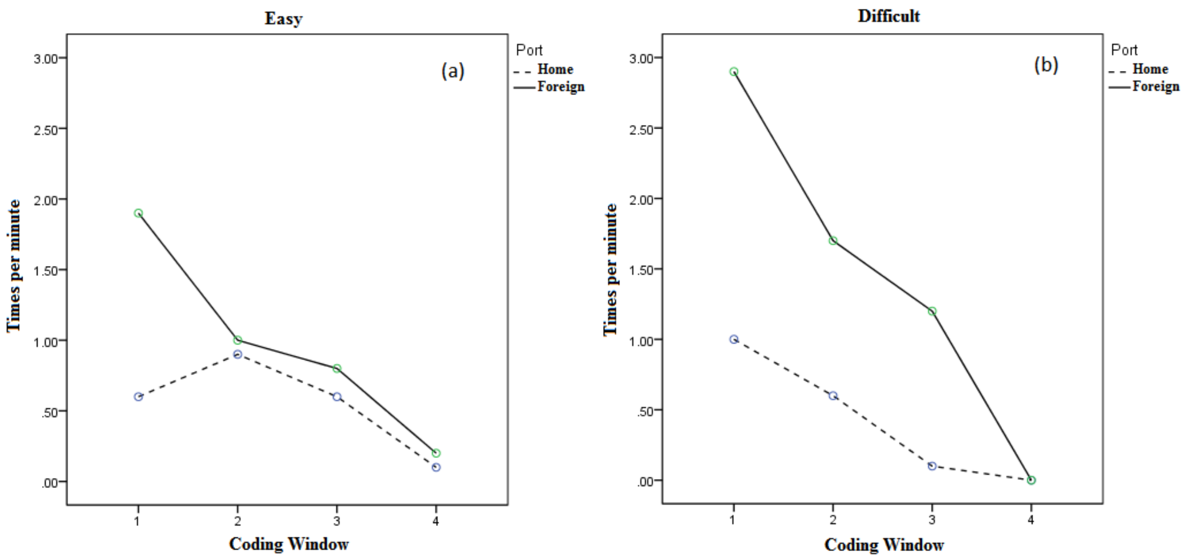



Even though a significant interaction was highlighted between the factors of difficulty and coding window (see figure 7(b) both lines in coding window 2), a comparison of the means (1 way ANOVA) using Bonferroni adjustment did not confirm such result. Having the capacity to reduce and maintain control of vessel speed is absolutely crucial to successfully perform a safe rotation in confined waters. The general trend was an increase in frequency of the speed checks in all the manoeuvres (CdWd_1 = 0.48, CdWd_2 = 2.85, CdWd_3 = 2.90, CdWd_4 = 4.00). To safely moor a vessel alongside a berth, pilots have to progressively reduce the speed of the ship in order to arrive alongside with minimal momentum. The lateral monitoring, while using the log of a ship (which provides information about the longitudinal and lateral speed), allows the pilot to calibrate the landing until the final touch on the fenders.

## Plan Check

The behavioural marker Plan Check recorded the number of times per minute that pilots were looking at their plan of how they would complete the manoeuvre. As reported in table B.6. in Appendix B, significant results were found both as main effects and as interactions between factors.

**Figure 8. Graphs of Plan Check**

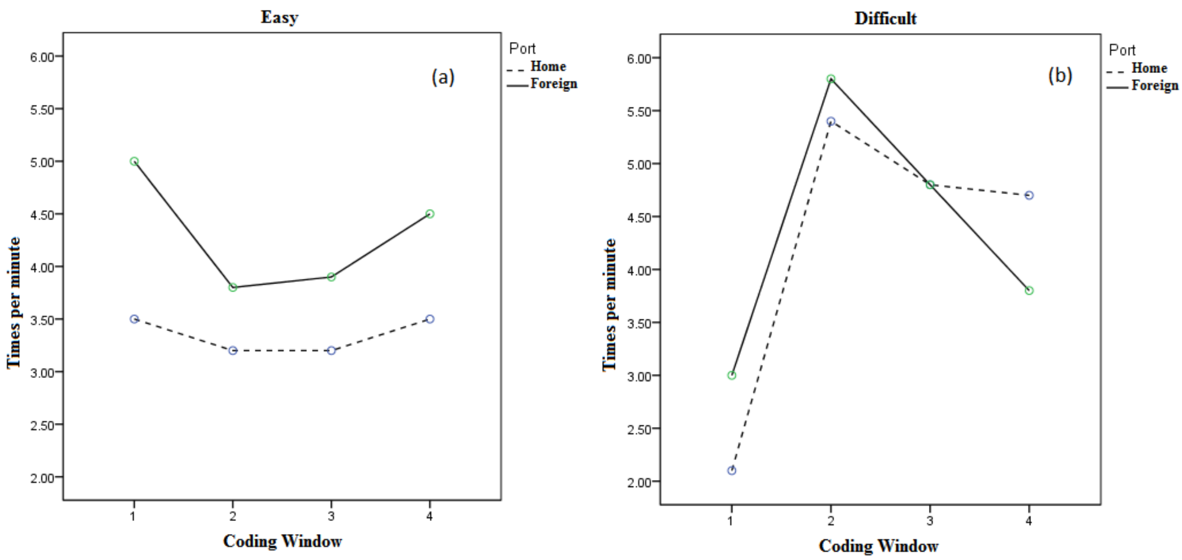



Both figures 8(a) and 8(b) highlight the main effect obtained on the factor port, with the foreign port (continuous line) consistently showing higher scores, as confirmed by both 3 way (Port_1 = 0.49, Port_2 = 1.21) and both 2 way ANOVAs (fixing the factor difficulty). In figures 8(a) and 8(b) a general decreasing trend is evident (CdWd_1 = 1.60, CdWd_2 = 1.05, CdWd_3 = 0.68, CdWd_4 = 0.08), showing higher scores at the very beginning of the manoeuvre and with the Closing significantly recording the lowest scores. It is evident that manoeuvring within an unfamiliar port forced pilots to refer more to the plan. Those plans, that were mainly navigational charts, were adopted especially during the exchange of information between Pilot and Captain or Pilot and VTS (using VHF), to locate elements of interest mentioned in those communications. Pilots referred more to those charts in the foreign port, to locate elements useful for ship positioning (transits, navigational aids, etc..) and to double check that they were following the intended plan.

### Pilot Orders

The behavioural marker Pilot Orders recorded the number of orders per minute the pilots gave, either to the Bridge personnel or, via radio, to the tugs. As reported in table B.7. in Appendix B, significant results were found both as main effects and as interactions between factors.

**Figure 9. Graphs of Pilot Orders**




The main effect found that the factor Coding Window in the 3 way ANOVA was not significant, when using Bonferroni adjustment to compare the means. The interaction between the factors of Difficulty and Coding window, is evident when comparing figures 9(a) and 9(b). This interaction was confirmed by 2 way ANOVA (considering the homeport and the foreign port separately and then, separately, the two levels of difficulty). The 2 way ANOVA showed a significant main effect of the factor port in the easy manoeuvres (Diff = 1) (Port_1 = 3.35, Port_2 = 4.30, as depicted in figure 9(a)). Considering only the homeport (Port = 1) a main effect on the factor difficulty was found, with higher scores for the difficult manoeuvre compared to the easy one (Diff_1 = 3.35, Diff_2 = 4.25). In the difficult manoeuvres (Diff = 2), represented in figure 9(b), the factor Coding Window recorded a significant main effect, with Approach_1 (CdWd_1 = 2.55) having the lowest scores, compared to Approach_2 (CdWd_2 = 5.60) and the Swing (CdWd_3 = 4.80). It is important to remember that the difficult manoeuvres (figure 9(a)) were performed not only through giving orders to the bridge personnel but also to the tugs. Pilots controlled the tugs giving them orders using a VHF radio. The use of tugs is particularly relevant when rotating and translating a ship. Thus it is not surprising that the highest rate of orders was achieved just before and during the swing phase (coding windows 2 and 3).

## Conclusions

This study demonstrates that pilots adopt different shiphandling techniques to respond to the specific challenges of each phase, which was evident through their adoption of different exploratory strategies for searching their environment.

Table 5 describes marine pilotage situational awareness according to Smith and Hancock's Perceptual Cycle model (K. Smith & Hancock, 1995). In this table we can identify the characteristics of internally managed principles and concepts of shiphandling (Pilotage Schema), in different conditions and situations (Phase Description). Within these schemata:

- Pilots are engaged in goal-directed action (i.e., they have particular shiphandling priorities at different phases of the manoeuvre. The pilot shifts from monitoring speed and direction, to focusing on position and momentum and then speed and lateral position on closing to the berth
- Pilot hold more generalised relevant information for that action, and this changes with the shiphandling priority. For example in the approach the focus on speed is associated with knowledge about maximum/minimum speeds necessary given the environmental and other conditions.
- Pilots direct their attention to very specific and more relevant sources of information (Exploration of the Object) to decide how to act. Continuing the example above, information on the ships speed is constantly monitored on the bridge.

The Perceptual Cycle model, also includes the active interaction between the subject and the environment through actions ("powerful behaviours"). These actions are constantly aiming to reduce the discrepancy between the perceived state (present SA), obtained through the information gathered, and the one aimed or desired based on one's own schemata (K. Smith & Hancock, 1995). In this study, the different use of those actions was highlighted by the different frequency of orders recorded in the different phases, levels of difficulty and familiarity of the manoeuvres. This study has, through the use of eye tracking and other technologies (i.e., simulators) been able to comprehensively map changes in the perceptual cycle of marine pilots and is the first example of a quantitative analysis in this domain.

**Table 5. Perceptual Cycle Model applied to results**

| **Phase** | **Approach** | **Approach** | **Swing** | **Closing** |
|---|---|---|---|---|
| **Coding Window** | **Coding Window 1 – Approach_1** | **Coding Window 2 – Approach_2** | **Coding Window 3 - Swing** | **Coding Window 4 - Closing** |
| **Phase Description** | Start of manoeuvre with vessel in a known position and at a desired speed, directed towards the berth | Vessel approaching the position were the swing is started | Completion of the 180° rotation of the vessel within the swinging basin spatial constrains | Final closing to the berth with suitable angle and speed in order to safely moor the vessel |
| **Shiphandling Priority (Schemata)** | At the starting of the manoeuvre, the vessel is still relatively distant from the expected area of rotation (swing basin). The priority at this stage is to monitor the direction and manage the speed at which the vessel is proceeding. The aim is to appreciate and counteract the effects of environmental forces on the vessel so to manoeuvre her correctly in order to reach the desired position where to safely start the swing. | At this stage, the speed should have already been consistently reduced in order to reach in a very controlled manner the exact position where the swing is expected to start. The priority is to position the vessel correctly, in order to have enough space to swing it around in the next phase.<br>Another goal is also to correctly reposition the tugs (when available) to have them applying the desired forces. | In this phase the vessel is completing the 180° turn. The rotation is obtained using tugs to apply transversal forces, while the main engine is used to control longitudinal momentum. In addition, the vessel continuously changes its relative angle to the environmental forces, developing new resultants that require to be managed / corrected to maintain the desired safe position. | The vessel, once stabilised her angular momentum at the end of the swing, has to be controlled in angle and speed, during her approach to the berth. Longitudinal and lateral speeds have to be reduced to a minimum before touching the fenders. |
| **Pilotage Knowledge (Schemata)** | Knowledge of heading values to be matched to aim in the right direction.<br>Knowledge of safety envelopes (max allowed speed / angle difference from expected heading) to ensure positive control of vessel.<br>Knowledge of effects of wind and current on the actual direction and speed of the vessel. | Knowledge of desired position to be reached in consideration of future effects that the vessel will experience during the swing.<br>Knowledge of use of tugs to obtain application of forces where required (when applicable). | Knowledge of the effects of environmental forces while the vessel is constantly changing angle.<br>Knowledge of use of tugs to obtain necessary forces in direction and intensity (when applicable). | Knowledge of heading values to be matched to have a suitable angle of approach with the berth.<br>Knowledge of safety envelopes (max allowed speed / angle difference from expected heading) to ensure positive control of vessel during landing.<br>Knowledge of effects of wind and current depending on the actual direction and speed of the vessel. |
| **Exploration of the object** | Use of visual cues against the forward mast of the ship, to perceive rotation and verify direction in the channel;<br>Use of dedicated equipment to check heading and rotation and speed.<br>Limited visual appreciation of vessel's speed looking at objects at the beam. | Use of external references (objects at the beam) to assess reduction of speed;<br>Use of external reference and Bridge equipment to verify exact vessel position before committing to the swing; | Use of external references and dedicated equipment to assess development of rotation.<br>Use of external reference and Bridge equipment to verify vessel position during rotation. | Comparison of vessel orientation against berth orientation (with the aim to reduce that angle to zero when landing).<br>Close monitoring of longitudinal and lateral speeds that have to remain within limits to allow a safe landing;<br>Close monitoring of vessel position against desired position alongside. |

**Future Applications and Added Value.**

From a methodological perspective, the results demonstrate how a full mission bridge simulator can be used to create different scenarios and situations, which induced different behavioural patterns. These patterns can be measured with unobtrusive tools, such as eye trackers, to record the locations and frequency of visual attention allocation.

An examination of the most relevant and useful source of information sought by pilots, depending on context and manoeuvring conditions, can assist designers and manufacturers to optimise equipment designs, and trainers to teach more efficient and appropriate shiphandling techniques. Being able to define and monitor meaningful behavioural markers can be used to improve evaluation of training outcomes, actual performance and, in the future, real time activities in normal working environments. Future studies involving gaze analysis (and therefore attention) specifically with reference to the source of information (electronic equipment, external visual aids..) preferred by pilots (Itoh et al., 1990) could offer an important insight regarding information resource management and shedding preferences once task demand begins to overcome pilot capabilities (M. S. Young et al., 2015). All of the above information will be useful in the reconstruction of accidents within simulated environments – an increasing practice in the maritime domain.

As this research was conducted in a simulated environment, future studies should consider the collection and comparison of similar data in a real working environment. For this to be achieved, it will be critical for data collection to be as resilient and unobtrusive as possible, in order not to distract or interfere with berthing operations. Should that data collection be possible, it could provide a better understanding of normal and abnormal, individual and group response levels, which could help to identify critical operations and levels of performance. Once those behavioural patterns were identified, they could be exploited as prodromal indicators of critical conditions or good performance. Beyond this, such data collection will provide a better appreciation of the realism of the simulated environment through a comparative analysis of the same behavioural markers.

Finally, we believe that this study could inform improvements in Australian national pilotage standards (ATC, 2008) around issues such as use of simulation facilities for training and continuing professional development of pilots. Improvements could include using our experimental protocol to update training standards, creating materials for CPD around the relationship between behavioural patterns and performance outcomes, and using these results as the basis for further studies, which could lead to modifications of shiphandling techniques.

Aasman, J., Mulder, G., & Mulder, L. J. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 29*(2), 161-170.

Aasman, J., Wijers, A., Mulder, G., & Mulder, L. (1988). Measuring mental fatigue in normal daily working routines. *Advances in Psychology, 52*, 117-137.

Abdi, H., & Williams, L. J. (2010). Tukey's honestly significant difference (HSD) test. *Encyclopedia of Research Design. Thousand Oaks, CA: Sage*, 1-5.

Abernethy, B. (1991). Visual search strategies and decision-making in sport. *International Journal of Sport Psychology, 22(3-4)*, 189-210.

Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. *Memory & cognition, 9*(4), 422-433. doi: 10.3758/bf03197568

Al-Diban, S., & Ifenthaler, D. (2011). Comparison of Two Analysis Approaches for Measuring Externalized Mental Models. *Educational Technology & Society, 14*(2), 16-30.

Alderton, T. (2004). *The global seafarer: living and working conditions in a globalized industry*: International Labour Organization.

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *The Quarterly journal of experimental psychology, 24*(2), 225-235.

AMSA. (2010). *Marine Orders Part 9: Health – Medical Fitness*.

AMSA. (2011). *Marine Orders Part 54: Coastal Pilotage*.

Anderson, J. R. (1982). Acquisition of Cognitive Skill. *Psychological Review, , v89* (n4 ), p369-406

Anderson, J. R. (1987). Skill Acquisition: Compilation of Weak-Method Problem Solutions. *Psychological Review, 94*(2), 192-210.

Aronson, J. L. (1997). Mental models and deduction. *American Behavioral Scientist, 40*(6), 782-797.

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review, 114*(3), 632-656.

ATC. (2008). *National marine guidance manual: guidelines for marine pilotage standards in Australia*. (ISBN 0 642 73653 7).

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The psychology of learning and motivation: Advances in research and theory, 2*, 89-195.

Babcock, P. (1976). Webster's third new international dictionary of the English language. *G. & C. Merriam Company, Springfield, MA, USA*.

Baldwin, C. L. (2003). Commentary - Neuroergonomics of mental workload new insights from the convergence of brain and behaviour in ergonomics research. *Theoretical Issues in Ergonomics Science, 4*(1-2), 132-141. doi: 10.1080/14639220210159807

Banks, A. P., & Millward, L. J. (2000). Running shared mental models as a distributed cognitive process. *British Journal of Psychology, 91*(4), 513-531.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*(4), 569.

Barrows, H. S. (1978). *An analysis of the clinical methods of medical students and physicians*: McMaster University.

Batmaz, I., & Öztürk, M. (2008). Using Pupil Diameter Changes for Measuring Mental Workload under Mental Processing. *Journal of Applied Sciences, 8*(1).

## References


Bechara, A. (2004). The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage. *Brain and Cognition, 55*(1), 30-40.

Bédard, J., Chi, M. T. H., Graham, L. E., & Shanteau, J. (1993). Expertise in auditing; Discussion. *Auditing, 12*, 21.

Bedny, G., & Meister, D. (1999). Theory of activity and situation awareness. *International Journal of cognitive ergonomics, 3*(1), 63-72.

Beilock, S. L., Bertenthal, B. I., McCoy, A. M., & Carr, T. H. (2004). Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review, 11*(2), 373-379.

Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of Experimental Psychology: Applied; Journal of Experimental Psychology: Applied, 8*(1), 6.

Beilock, S. L., Wierenga, S. A., & Carr, T. H. (2002). Expertise, attention, and memory in sensorimotor skill execution: Impact of novel task constraints on dual-task performance and episodic memory. *The Quarterly Journal of Experimental Psychology: Section A, 55*(4), 1211-1240.

Benedict, K. G., Michael; Kirchhoff, Matthias; Fischer, Sandro; Schaub, Michèle; Wismar, Hochschule;. (2012). Application of fast time manoeuvring simulation for ship handling in simulator training and on-board *INSLC 17 - International Navigation Simulator Lecturers' Conference*.

Benner, P. (1982). From novice to expert. *American Journal of Nursing*(March), 402-407.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., . . . Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine, 78*(5), B231-B244.

Berntson, G. G. (1997). Heart rate variability: origins, methods and interpretive caveats. *Psychophysiology, 34*, 623-648.

Beyea, S. C. (2005). High reliability theory and highly reliable organizations. *AORN Journal, 81*(6), 1319-1322. doi: 10.1016/S0001-2092(06)60397-9

Biggs, J., & Tang, C. (2010). *Applying constructive alignment to outcomes-based teaching and learning.* Paper presented at the Training material for "quality teaching for learning in higher education" workshop for master trainers, Ministry of Higher Education, Kuala Lumpur.

Biggs, S. F., & Wild, J. J. (1985). An Investigation of Auditor Judgment in Analytical Review. *The Accounting Review, 60*(4), 607-633. doi: 10.2307/247458

Billings, C. E. (1991). Human-centered aircraft automation: A concept and guidelines: NASA.

Bjørneseth, F. B., Clarke, L., Dunlop, M., & Komandur, S. (2014). *Towards an understanding of operator focus using eye-tracking in safety-critical maritime settings.* Paper presented at the International Conference on Human Factors in Ship Design & Operation.

Blix, A., Stromme, S., & Ursin, H. (1974). Additional heart rate--an indicator of psychological activation. *Aerospace Medicine, 45*(11), 1219.

Bloom, B. S. (1985). Generalizations about talent development. *Developing talent in young people*, 507-549.

Bloom, B. S., & Sosniak, L. A. (1985). *Developing talent in young people*: Ballantine Books.

Bolstad, B. M., Irizarry, R. A., Åstrand, M., & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics, 19*(2), 185-193.

Bolstad, C. A., & Hess, T. M. (1995). *Situation awareness and older workers.* Paper presented at the Experimental Analysis and Measurement of Situation Awareness Conference, Daytona Beach, FL.

Boose, J. H. (1986). Expertise transfer for expert system design. *System Research, 152*, 5.9.

Bordage, G., & Zacks, R. (1984). The structure of medical knowledge in the memories of medical students and general practitioners: categories and prototypes. *Medical Education, 18*(6), 406-416.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews, 44*, 58-75.

Bornstein, B. H., Christine Emler, A., & Chapman, G. B. (1999). Rationality in medical treatment decisions: is there a sunk-cost effect? *Social Science & Medicine, 49*(2), 215-222. doi: 10.1016/S0277-9536(99)00117-3

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*: sage.

Bramer, M. A., Bramer, M., & Bramer, M. (1985). *Research and Development in Expert Systems: Proceedings of the Fourth Technical Conference of the British Computer Society Specialist Group on Expert Systems, University of Warwick, 18-20 December 1984* (Vol. 4): Cambridge University Press.

Brand, M., Labudda, K., & Markowitsch, H. J. (2006). Neuropsychological correlates of decision-making in ambiguous and risky situations. *Neural Networks, 19*(8), 1266-1276.

Breton, R., & Rousseau, R. (2001). Situation Awareness: A review of the concept and its measurement. *DRDC Valcartier TR, 220*.

Brezovic, C. P., Klein, G. A., & Thordsen, M. (1990). Decision making in armored platoon command: DTIC Document.

Brookhuis, K. A., & De Waard, D. (1993). The use of psychophysiology to assess driver status. *Ergonomics, 36*(9), 1099-1110.

Brookhuis, K. A., Waard, D. d., & Fairclough, S. H. (2003). Criteria for driver impairment. *Ergonomics, 46*(5), 433-445.

Brooks, B. P., Coltman, T. C., & Yang, M. (2016). Technological Innovation in the Maritime Industry: The Case of Remote Pilotage and Enhanced Navigational Assistance. *Journal of Navigation, 69*(4), 777-793.

Bruzzone, A., Longo, F., Nicoletti, L., & Diaz, R. (2012). *Traffic controllers and ships pilots training in marine ports environments*. Paper presented at the Proceedings of the 2012 Symposium on Emerging Applications of M&S in Industry and Academia Symposium, Orlando, Florida.

Bryan, W. L., & Harter, N. (1897). Studies in the physiology and psychology of the telegraphic language. *Psychological Review, 4*(1), 27-53.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review, 6*(4), 345-375.

Bürki-Cohen, J., Soja, N. N., & Longridge, T. (1998). *Simulator fidelity requirements: the case of platform motion.* Paper presented at the International Training and Education Conference and Exhibition, Lausanne, Switzerland.

Cain, B. (2007). A review of the mental workload literature: DTIC Document.

Calderwood, R., Klein, G. A., & Crandall, B. W. (1988). Time pressure, skill, and move quality in chess. *The American Journal of Psychology*, 481-493.

Camerer, C. F., & Johnson, E. J. (1997). The process-performance paradox in expert judgment: How can experts know so much and predict so badly? *Research on judgment and decision making: Currents, connections, and controversies*, 342.

Carswell, C. M., Clarke, D., & Seales, W. B. (2005). Assessing mental workload during laparoscopic surgery. *Surgical innovation, 12*(1), 80-90.

References


Casali, J. G., & Wierwille, W. W. (1984). On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues. *Ergonomics, 27*(10), 1033-1050. doi: 10.1080/00140138408963584

Cellier, J. M., Eyrolle, H., & Marine, C. (1997). Expertise in dynamic environments. *Ergonomics, 40*(1), 28-50. doi: 10.1080/001401397188350

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*: The Guilford Press.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory, 2*(6), 641-653. doi: 10.1037/h0031185

Charness, N., Krampe, R., & Mayr, U. (1996). The role of practice and coaching in entrepreneurial skill domains: An international comparison of life-span chess skill acquisition. *The road to excellence: the acquisition of expert performance in the arts and sciences, sports, and games (Mahwah, NJ, Erlbaum)*, 51-80.

Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology, 19*(2), 151-165.

Chartrand, J.-P., Peretz, I., & Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain Research, 1220*, 191-198.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. *Visual information processing*, 215-281.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices*. *Cognitive Science, 5*(2), 121-152. doi: 10.1207/s15516709cog0502_2

Chi, M. T. H., Glaser, R., & Rees, E. (1981). Expertise in Problem Solving: Pittsburgh University Pa Learning Research And Development Center.

Chi, M. T. H., Glaser, R. E., & Farr, M. J. (1988). *The nature of expertise*: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H., Ohlsson, S., & Holyoak, K. (2005). Complex declarative learning. *The Cambridge handbook of thinking and reasoning*, 371-399.

Chi, M. T. H. E. G., Robert (Ed); Farr, M. J. (Ed) (1988). *The nature of expertise*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*(3), 257-273. doi: 10.1016/S0022-5371(79)90146-4

CMPA. (2017). Marine Pilotage in Canada: A Cost Benefit Analysis. In C. M. P. Association (Ed.): Transportation Economics & Management Systems, Inc.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement, 20*(1), 37-46.

Cohen, M. S. (1993). Three paradigms for viewing decision biases. *Decision making in action: Models and methods, 1*, 36-50.

Colle, H. A., & Reid, G. B. (1999). Double trade-off curves with different cognitive processing combinations: Testing the cancellation axiom of mental workload measurement theory. *Human Factors, 41*(1), 35-35.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: an elementary introduction*. Oxford, England: Prentice-Hall.

Council, N. R., & Piloting, C. o. A. i. N. (1994). *Minding the helm: marine navigation and piloting*: National Academy Press.

Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant Capacity in an Immediate Serial-Recall Task A Logical Sequel to Miller (1956). *Psychological Science, 15*(9), 634-640.

Crandall, B., & Calderwood, R. (1989). Clinical assessment skills of experienced neonatal intensive care nurses. *Final report, Klein Associates Inc., OH. Prepared under contract, 1*, R43.

Crandall, B., Kyne, M., Militello, L., & Klein, G. (1992). Describing expertise in one-on-one instruction. *Fairborn, OH: Klein Associates*.

Cranley, L., Doran, D. M., Tourangeau, A. E., Kushniruk, A., & Nagle, L. (2009). Nurses' uncertainty in decision-making: a literature review. *Worldviews on Evidence-Based Nursing, 6*(1), 3-15. doi: 10.1111/j.1741-6787.2008.00138.x

Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life*: Basic Books.

Currib, L. (1969). The Perception of Danger in a Simulated Driving Task. *Ergonomics, 12*(6), 841-849. doi: 10.1080/00140136908931101

Curry, R., Jex, H., Levison, W., & Stassen, H. (1979). Final report of control engineering group. *Mental workload: Its theory and measurement*, 235-252.

Dahlström, N. (2008). Pilot training in our time - use of flight training devices and simulators. [Article]. *Mokomųjų skraidymo prietaisų bei treniruoklių naudojimas orlaivių pilotų mokymui šiais laikais., 12*(1), 22-27.

Dahlstrom, N., Dekker, S., Van Winsen, R., & Nyce, J. (2009). Fidelity and validity of simulator training. *Theoretical Issues in Ergonomics Science, 10*(4), 305-314.

Damasio, A. R., Everitt, B., & Bishop, D. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex [and discussion]. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 351*(1346), 1413-1420.

Dann, P. (2012). *Development and Validation of Situational Awareness Behavioural Markers for Maritime Pilotage.* (Masters of Human Factors), University of South Australia.

Dawes, R. (1996). *House of cards*: Free Press.

De Groot, A. D. (1978). *Thought and choice in chess* (Vol. 4): Walter de Gruyter.

de Vries, L. (2015). Success factors for navigational assistance: a complementary ship-shore perspective. *Proceedings of the human factors and ergonomics society Europe*, 175-186.

De Waard, D. (1996). *The measurement of drivers' mental workload*. Haren, The Netherlands: Groningen University, Traffic Research Center.

Di Nocera, F., Mastrangelo, S., Colonna, S. P., Steinhage, A., Baldauf, M., & Kataria, A. (2016). Mental workload assessment using eye-tracking glasses in a simulated maritime scenario. *Proceedings of the human factors and ergonomics society Europe*.

Di Stasi, L. L., Contreras, D., Cándido, A., Cañas, J., & Catena, A. (2011). Behavioral and eye-movement measures to track improvements in driving skills of vulnerable road users: First-time motorcycle riders. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*(1), 26-35.

Di Stasi, L. L., Diaz Piedra, C., Suárez, J., McCamy, M. B., Martinez Conde, S., Roca Dorda, J., & Catena, A. (2015). Task complexity modulates pilot electroencephalographic activity during real flights. *Psychophysiology, 52*(7), 951-956.

Dijksterhuis, A., & Nordgren, L. F. (2006). A Theory of Unconscious Thought. *Perspectives on Psychological Science, 1*(2), 95-109. doi: 10.2307/40212159

Dijkstra, K., MacMahon, C., & Misirlisoy, M. (2008). The effects of golf expertise and presentation modality on memory for golf and everyday items. *Acta Psychologica, 128*(2), 298-303.

DNV_GL_AS. (2014). Standard DNVGL-ST-0033:2014-08 Maritime simulator systems.

Doane, S. M., Pellegrino, J. W., & Klatzky, R. L. (1990). Expertise in a Computer Operating System: Conceptualization and Performance. [Article]. *Human-Computer Interaction, 5*(2/3), 267.

Doane, S. M., & Sohn, Y. W. (2000). ADAPT: A Predictive Cognitive Model of User Visual Attention and Action Planning. *User Modeling and User-Adapted Interaction, 10*(1), 1-45. doi: 10.1023/a:1008311003128

Dreyfus, H. (1972). *What Computer Still Can't Do: A Critique of Artificial Reason*: MIT Press, original edition published in.

Dreyfus, H. L., Dreyfus, S. E., & Zadeh, L. A. (1987). Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer. *IEEE Expert, 2*(2), 110-111. doi: 10.1109/mex.1987.4307079

Dreyfus, S. E., & Dreyfus, H. L. (1980). A five-stage model of the mental activities involved in directed skill acquisition: DTIC Document.

Duchowski, A. T. (2007). *Eye tracking methodology: theory and practice*: Springer.

Duffy, L. J., Baluch, B., & Ericsson, K. A. (2004). Dart performance as a function of facets of practice amongst professional and amateur men ana women players. *Int. J. Sport Psychol, 35*, 232-245.

Dunn, B. D., Dalgleish, T., & Lawrence, A. D. (2006). The somatic marker hypothesis: A critical evaluation. *Neuroscience and biobehavioral reviews, 30*(2), 239.

Durso, F. T., & Gronlund, S. D. (1999). Situation awareness. *Handbook of applied cognition*, 283-314.

Dussault, C., Jouanin, J.-C., Philippe, M., & Guezennec, C.-Y. (2005). EEG and ECG changes during simulator operation reflect mental workload and vigilance. *Aviation, Space, and Environmental Medicine, 76*(4), 344-351.

Ebbinghaus, H. (1967). *Memory: A contribution to experimental psychology*: Dover New York.

Egan, D. E. (1979). Chunking in recall of symbolic drawings. *Memory & cognition, 7*(2), 149-158. doi: 10.3758/bf03197595

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. L. (1991). Workload assessment in multi-task environments. *Multiple-task performance*, 207-216.

Eggleton, I. R. (1982). Intuitive time-series extrapolation. *Journal of Accounting Research*, 68-102.

Eisenstadt, M., & Kareev, Y. (1975). Aspects of human problem solving: The use of internal representations. *Explorations in cognition*, 308-346.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning* (Vol. 2): Harvard University Press Cambridge, MA.

Endsley, M. (1988). A construct and its measurement: The functioning and evaluation of pilot situation awareness (No. NOR DOC 88-30). *Hawthorne, CA: Northrop Corporation*.

Endsley, M. R. (1988). *Design and evaluation for situation awareness enhancement.* Paper presented at the Human Factors Society, Annual Meeting, 32 nd, Anaheim, CA.

Endsley, M. R. (1995). Direct measurement of situation awareness in simulations of dynamic systems: validity and use of SAGAT *Experimental Analysis and Measurement of Situation Awareness*, 107-113.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 32-64.

Endsley, M. R. (1999). *Situation awareness and human error: Designing to support human performance.* Paper presented at the Proceedings of the High Consequence Systems Surety Conference.

Endsley, M. R. (2006). Expertise and Situation Awareness *Cambridge Handbook of Expertise and Expert Performance*: Cambridge University Press.

Endsley, M. R., & Bolstad, C. A. (1994). Individual differences in pilot situation awareness. *The International Journal of Aviation Psychology, 4*(3), 241-264.

Endsley, M. R., & Garland, D. (2000). Theoretical underpinnings of situation awareness: A critical review. *Situation awareness analysis and measurement*, 3-32.

Endsley, M. R., & Garland, D. J. (2000). *Situation Awareness: Analysis and Measurement*: Lawrence Erlbaum Associates.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*(2), 381-394.
Engle, R. W., & Bukstel, L. (1978). Memory processes among bridge players of differing expertise. *The American Journal of Psychology*, 673-689.
Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*, 709-709.
Eric Farmer, A. B. (2003). Review of Workload Measurement, Analysis and Interpretation Methods *European Organisation For The Safety Of Air Navigation* (Report).
Ericsson, K., & Simon, H. (1993). Protocol analysis: Verbal reports as data (rev. ed.) MIT Press. *Cambridge, MA*.
Ericsson, K. A. (1990). Peak performance and age: An examination of peak performance in sports. *Successful aging: Perspectives from the behavioral sciences*, 164-195.
Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. *The road to excellence: the acquisition of expert performance in the arts and sciences, sports and games. Mahwah (NJ): Lawrence Erlbaum Associates*, 1-50.
Ericsson, K. A. (2001). The path to expert golf performance: Insights from the masters on how to improve performance by deliberate practice. *Optimising performance in golf*, 1-57.
Ericsson, K. A. (2002). Attaining Excellence Through Deliberate Practice: Insights from the Study of Expert Performance *Teaching and Learning: The Essential Readings* (pp. 4-37): Blackwell Pub.
Ericsson, K. A. (2003). The acquisition of expert performance as problem solving. *The psychology of problem solving*, 31-83.
Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*(10), S70-S81.
Ericsson, K. A. (2004). Uncovering the structure of a memorist's superior "basic" memory capacity. *Cognitive Psychology, 49*(3), 191-237. doi: 10.1016/j.cogpsych.2004.02.001
Ericsson, K. A. (2006a). Cambridge Handbook of Expertise and Expert Performance. *Cambridge Universal Press*.
Ericsson, K. A. (2006b). The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance *Cambridge Handbook of Expertise and Expert Performance*: Cambridge University Press.
Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (2006). *The Cambridge handbook of expertise and expert performance*: Cambridge University Press.
Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245.
Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406. doi: 10.1037/0033-295x.84.2.127.
Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273-273.
Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: An introduction. *Toward a general theory of expertise: Prospects and limits*, 1-38.
Ericsson, K. A., Starkes, J., & Ericsson, K. (2003). Development of elite performance and deliberate practice. *Expert performance in sports: Advances in research on sport expertise*, 49-83.
Ericsson, K. A. K., Walter. (2000). Shortcomings of generic retrieval structures with slots of the type that Gobet (1993) proposed and modelled. [Article]. *British Journal of Psychology, 91*(4), 571.

References

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol., 59*, 255-278.
Evans, J. S. B. (2009). Introspection, confabulation, and dual-process theory. *Behavioral and Brain Sciences, 32*(02), 142-143.
Faghih-Roohi, S., Xie, M., & Ng, K. M. (2014). Accident risk assessment in marine transportation via Markov modelling and Markov Chain Monte Carlo simulation. *Ocean Engineering, 91*, 363-370.
Falkmer, T., & Gregersen, N. P. (2005). A comparison of eye movement behavior of inexperienced and experienced drivers in real traffic environments. *Optometry & Vision Science, 82*(8), 732-739.
Fallesen, J. J., & Pounds, J. (2001). Identifying and testing a naturalistic approach for cognitive skill training. *Linking Expertise and Naturalistic Decision Making. Mahwah, NJ: Lawrence Erlbaum Associates*, 55-70.
Farmer, E. W., Berman, J. V., & Fletcher, Y. L. (1986). Evidence for a visuo-spatial scratch-pad in working memory. *The Quarterly journal of experimental psychology, 38*(4), 675-688.
Farrington-Darby, T., & Wilson, J. R. (2006). The nature of expertise: A review. *Applied Ergonomics, 37*(1), 17-32. doi: 10.1016/j.apergo.2005.09.001
Feil, A., & Mestre, J. P. (2010). Change Blindness as a Means of Studying Expertise in Physics. *Journal of the Learning Sciences, 19*(4), 480-505. doi: 10.1080/10508406.2010.505139
Feltovich, P. J., Johnson, P. E., Moller, J. H., & Swanson, D. B. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. *Readings in medical artificial intelligence*, 275-319.
Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. *The Cambridge handbook of expertise and expert performance*, 41-67.
Fiol, C. M., & Huff, A. S. (1992). Maps for managers: where are we? Where do we go from here? . [Article]. *Journal of Management Studies, 29*(3), 267-285.
Fisher, D., Pradhan, A., Pollatsek, A., & Knodler Jr, M. (2007). Empirical evaluation of hazard anticipation behaviors in the field and on driving simulator using eye tracker. *Transportation Research Record: Journal of the Transportation Research Board*(2018), 80-86.
Fitts, P. M. (1966). Cognitive Aspects of Information Processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology, 71*(6), 849.
Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Oxford, England: Brooks/Cole.
Flin, R., Slaven, G., & Stewart, K. (1996). Emergency decision making in the offshore oil and gas industry. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 38*(2), 262-277.
Forsman, F., Sjors, A., Dahlman, J., Falkmer, T., & Lee, H. C. (2012). Eye Tracking During High Speed Navigation at Sea. *Journal of Transportation Technologies, Vol.02No.03*, 8. doi: 10.4236/jtts.2012.23030
Gaillard, A. W. K. (1993). Comparing the concepts of mental load and stress. *Ergonomics, 36*(9), 991-1005. doi: 10.1080/00140139308967972
Galton, S. F. (1869). *Hereditary genius*: Macmillan and Company.
Garrett, S. K., Caldwell, B. S., Harris, E. C., & Gonzalez, M. C. (2009). Six dimensions of expertise: A more comprehensive definition of cognitive expertise for team coordination. *Theoretical Issues in Ergonomics Science, 10*(2), 93-105.
Gawel, R. (1997). The use of language by trained and untrained experienced wine tasters. *Journal of Sensory Studies, 12*(4), 267-284. doi: 10.1111 /j.1745-459X.1997. tb00067.x

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise Differences in the Comprehension of Visualizations: a Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educational Psychology Review, 23*(4), 523-552.
Gekara, V. O., Bloor, M., & Sampson, H. (2011). Computer-based assessment in safety-critical industries: the case of shipping. *Journal of Vocational Education & Training, 63*(1), 87.
Gentner, D. (1988). Metaphor as Structure Mapping: The Relational Shift. *Child Development, 59*(1), 47-59. doi: 10.2307/1130388
Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., & Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 40*(1), 79-91.
Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral Cortex, 7*(4), 374-385.
Ghosh, S. (2017). Can authentic assessment find its place in seafarer education and training? *Australian Journal of Maritime & Ocean Affairs*, 1-14. doi: 10.1080/18366503.2017.1320828
Ghosh, S., Bowles, M., Ranmuthugala, D., & Brooks, B. (2014). Reviewing seafarer assessment methods to determine the need for authentic assessment. *Australian Journal of Maritime & Ocean Affairs, 6*(1), 49-63.
Ghosh, S., Bowles, M., Ranmuthugala, D., & Brooks, B. (2016). Authentic assessment in seafarer education: using literature review to investigate its validity and reliability through rubrics. *WMU Journal of Maritime Affairs, 15*(2), 317-336.
Gilbert, D. T. (1999). What the mind’s not. *Dual-process theories in social psychology*, 3-11.
Gobet, F., & Chassy, P. (2009). Expertise and intuition: A tale of three theories. *Minds and Machines, 19*(2), 151-180.
Gobet, F. S., Herbert. (1996). Templates in Chess Memory: A Mechanism for Recalling Several Boards. *Cognitive Psychology, 31*(1), 1-40. doi: 10.1006/cogp.1996.0011
Goodfellow, M. (2008). *Are Crew Competencies Declining? A cause of concern for us all!* Paper presented at the Coasts and Ports Australasian Conference, Newcastle, Australia.
Gopher, D., & Braune, R. (1984). On the Psychophysics of Workload: Why Bother with Subjective Measures? *Human Factors: The Journal of the Human Factors and Ergonomics Society, 26*(5), 519-532. doi: 10.1177/001872088402600504
Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman & J. P. E. Thomas (Eds.), *Handbook of perception and human performance, Vol. 2. Cognitive processes and performance* (Vol. Vol. 2). Oxford, England: John Wiley.
Gould, K. S., Røed, B. K., Saus, E.-R., Koefoed, V. F., Bridger, R. S., & Moen, B. E. (2009). Effects of navigation method on workload and performance in simulated high-speed ship navigation. *Applied Ergonomics, 40*(1), 103-114.
Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6*(12), 517-523.
Greenwood, J., & King, M. (1995). Some surprising similarities in the clinical reasoning of "expert" and "novice" orthopaedic nurses: report of a study using verbal protocols and protocol analyses. *Journal of Advanced Nursing, 22*(5), 907-913. doi: 10.1111/j.1365-2648.1995.tb02642.x
Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., & St. John, M. (2008). *The red-line of workload: Theory, research, and design.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Grossman, P. (1983). Respiration, stress, and cardiovascular function. *Psychophysiology, 20*(3), 284-300.

Gruber, H., Degner, S., & Lehmann, A. C. (2004). Why do some commit themselves in deliberate practice for many years–and so many do not? Understanding the development of professionalism in music. *Current issues in adult learning and motivation*, 222-235.

Gundel, A., & Wilson, G. F. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain topography, 5*(1), 17-25.

Gustin, W. C., & Bloom, B. (1985). The development of exceptional research mathematicians. *Developing talent in young people*, 270-331.

Hagemann, K. (2008). *The alpha band as an electrophysiological indicator for internalized attention and high mental workload in real traffic driving (Ph. D. thesis).* University of Düsseldorf, Germany. Retrieved from https://docserv.uni-duesseldorf.de/servlets/DerivateServlet/Derivate-8802/Dissertation_Konrad_Hagemann_pdfa.pdf

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*(5827), 998-1002.

Hambrick, D. (2002). Effects of Domain Knowledge, Working Memory Capacity, and Age on Cognitive Performance: An Investigation of the Knowledge-Is-Power Hypothesis. *Cognitive Psychology, 44*(4), 339-387. doi: 10.1006/cogp.2001.0769

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*: Oxford University Press on Demand.

Hancock, P., & Chignell, M. H. (1986). Toward a Theory of Mental Work Load: Stress and Adaptability in Human-Machine Systems. *Proc. IEEE SMC 1986*, 378-383.

Hancock, P. A., & Meshkati, N. (1988). *Human Mental Workload*: North-Holland.

Hansson, M., & Jönsson, P. (2006). Estimation of HRV spectrogram using multiple window methods focussing on the high frequency power. *Medical engineering & physics, 28*(8), 749-761.

Hareide, O. S., Ostnes, R., & Mjelde, F. V. (2016). *Understanding the Eye of the Navigator.* Paper presented at the European Navigation Conference Proceedings.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload, 1*, 139-183.

Hartman, B. O., & McKenzie, R. E. (1979). Survey of methods to assess workload: DTIC Document.

Hatano, G., Miyake, Y., & Binks, M. G. (1977). Performance of expert abacus operators. *Cognition, 5*(1), 47-55. doi: 10.1016/0010-0277(77)90016-6

Hayes, J. R. (1989). *The complete problem solver (2nd ed.)*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Helsen, W. F., Starkes, J. L., & Hodges, N. J. (1998). Team sports and the theory of deliberate practice. *Journal of Sport & Exercise Psychology; Journal of Sport & Exercise Psychology*.

Hermens, F., Flin, R., & Ahmed, I. (2013). Eye movements in surgery: A literature review. *Journal of Eye Movement Research, 6*(4).

Hetherington, C., Flin, R., & Mearns, K. (2006). Safety in shipping: The human element. *Journal of Safety Research, 37*(4), 401-411. doi: 10.1016/j.jsr.2006.04.007

Hinsley, D. A., Hayes, J. R., & Simon, H. A. (1977). From words to equations: Meaning and representation in algebra word problems. *Cognitive processes in comprehension*, 89-106.

Hinsz, V. B. (1995). Mental Models of Groups as Social Systems Considerations of Specification and Assessment. *Small Group Research, 26*(2), 200-233.

Hodges, N. J., & Starkes, J. (1996). Wrestling with the nature expertise: a sport specific test of Ericsson, Krampe and Tesch-Römer's (1993) theory of" deliberate practice". *International Journal of Sport Psychology, 27*(4), 400-424.

Hoffman, R. R. (1992). *The psychology of expertise: Cognitive research and empirical AI.* Paper presented at the This volume is based on papers presented at a conference entitled" Expert Systems and the Psychology of Expertise," held at Adelphi University, Garden City, NY, May 5, 1989.

Hoffman, R. R., Crandall, B., & Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 40*(2), 254-276.

Hoffman, R. R., Shadbolt, N. R., Burton, A. M., & Klein, G. (1995). Eliciting knowledge from experts: A methodological analysis. *Organizational Behavior and Human Decision Processes, 62*(2), 129-158.

Hollands, J. G., & Wickens, C. D. (1999). *Engineering psychology and human performance*: Prentice Hall New Jersey.

Holyoak, K. J. (1991). Symbolic connectionism: toward third-generation theories of expertise. *Toward a general theory of expertise: Prospects and limits*, 301.

Hontvedt, M. (2015). Professional vision in simulated environments—Examining professional maritime pilots' performance of work tasks in a full-mission ship simulator. *Learning, Culture and Social Interaction*.

Hontvedt, M., & Arnseth, H. C. (2013). On the bridge to learn: Analysing the social organization of nautical instruction in a ship simulator. *International Journal of Computer-Supported Collaborative Learning, 8*(1), 89-112.

Horrey, W. J., & Wickens, C. D. (2003). *Multiple resource modeling of task interference in vehicle control, hazard awareness and in-vehicle task performance.* Paper presented at the Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design.

Horswill, M. S., & McKenna, F. P. (2004). Drivers' hazard perception ability: Situation awareness on the road. *A cognitive approach to situation awareness: Theory and application*, 155-175.

Hreniuc, V., & Batrinca, G. (2014). A Pleading for Ship Manned Models as a "Physical" Simulator in the Ship Handling Training Process. *Procedia Engineering, 69*, 1410-1419.

Hsu, W.-K. K. (2015). Assessing the Safety Factors of Ship Berthing Operations. *Journal of Navigation, 68*(03), 576-588. doi: 10.1017/s0373463314000861

Huey, B. M., & Wickens, C. D. (1993). *Workload transition: Implications for individual and team performance*: National Academies Press.

IBM_Corp. (2010). IBM SPSS Statistics for Windows (Version 2010). NY: IBM Corp.

ICS. (2007). *Bridge procedures guide 2007 - International Chamber of Shipping*: London : Marisec, c2007. 4th ed.

ICS. (2016). Shipping and World Trade, from http://www.ics-shipping.org/shipping-facts/shipping-and-world-trade; http://www.ics-shipping.org/shipping-facts/shipping-and-world-trade/number-and-nationality-of-world's-seafarers

IMO. (1978). *International Convention on Standards of Training, Certification and Watchkeeping of Seafarers - As Amended*. London, England.

IMO. (2001). *IMO Standard Marine Communication Phrases: Resolution A.918(22)*: International Maritime Organization.

IMO. (2006). *Adoption of the revised performance standards for electronic chart display and information systems (ECDIS)*. London, England.

IMO. (2013). *Development of an E-navigation strategy implementation plan*. London, England.

Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., & Bailey, B. P. (2005). *Towards an index of opportunity: understanding changes in mental workload during task execution.*

Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
Itoh, Y., Hayashi, Y., Tsukui, I., & Saito, S. (1990). The ergonomic evaluation of eye movement and mental workload in aircraft pilots. *Ergonomics, 33*(6), 719-732.
Jansma, J. M., Ramsey, N. F., Slagter, H. A., & Kahn, R. S. (2001). Functional Anatomical Correlates of Controlled and Automatic Processing. *Journal of Cognitive Neuroscience, 13*(6), 730-743. doi: 10.1162/08989290152541403
Jeannot, E. (2000). Situation awareness: Synthesis of literature search. *EEC Note No*.
Jeffries, R., Turner, A. A., Polson, P. G., & Atwood, M. E. (1981). The processes involved in designing software. *Cognitive skills and their acquisition, 255*, 283.
Jex, H. R. (1988). Measuring Mental Workload: Problems, Progress, and Promises. In A. H. Peter & M. Najmedin (Eds.), *Advances in Psychology* (Vol. Volume 52, pp. 5-39): North-Holland.
Ji, Q., Zhu, Z., & Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE transactions on vehicular technology, 53*(4), 1052-1068.
Jiang, W., Hayano, J., Coleman, E., Hanson, M. W., Frid, D. J., O'Conno, C., . . . Blumenthal, J. A. (1993). Relation of cardiovascular responses to mental stress and cardiac vagal activity in coronary artery disease. *The American journal of cardiology, 72*(7), 551-554.
Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*: Harvard University Press.
Johnson, P. E., Duran, A. S., Hassebrock, F., Moller, J., Prietula, M., Feltovich, P. J., & Swanson, D. B. (1981). Expertise and Error in Diagnostic Reasoning*. *Cognitive Science, 5*(3), 235-283. doi: 10.1207/s15516709cog0503_3
Jorna, P. G. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology, 34*(2), 237-257.
Jou, Y.-T., Yenn, T.-C., Lin, C. J., Yang, C.-W., & Chiang, C.-C. (2009). Evaluation of operators' mental workload of human–system interface automation in the advanced nuclear power plants. *Nuclear Engineering and Design, 239*(11), 2537-2542.
Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*: Cambridge University Press.
Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition, 11*(2), 123-141. doi: 10.1016/0010-0277(82)90022-1
Kalinowski, A. G., & Bloom, B. (1985). The development of Olympic swimmers. *Developing talent in young people*, 139-192.
Kalsbeek, J., & Ettema, J. (1963). Scored regularity of the heart rate pattern and the measurement of perceptual or mental load. *Ergonomics, 6*(3), 306-307.
Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational Implications of Expertise Reversal Effects in Learning and Performance of Complex Cognitive and Sensorimotor Skills. *Educational Psychology Review, 24*(2), 313-337.
Keller, F. S. (1958). The phantom plateau. *Journal of the Experimental Analysis of Behavior, 1*(1), 1.
Keren, G. (1987). Facing uncertainty in the game of bridge: A calibration study. *Organizational Behavior and Human Decision Processes, 39*(1), 98-114.
Kipp, M. (2010). Anvil: The video annotation research tool. *Handbook of Corpus Phonology. Oxford University Press, Oxford (to appear, 2011)*.
Kitamura, K., Murai, K., Hayashi, Y., Fujita, T., & Maenaka, K. (2014). *Measurement and analysis of marine pilot's performance using a large model sensor.* Paper presented at the Consumer Electronics (GCCE), 2014 IEEE 3rd Global Conference on.
Kitson, A., Dawes, M., Davies, P. T., Gray, A., Mant, J., Seers, K., & Snowball, R. (1999). *Evidence-based practice: a primer for health care professionals*: Churchill Livingstone Edinburgh.

Klein, G. (1997). The current status of the naturalistic decision making framework. *Decision making under stress- Emerging themes and applications(A 99-12526 01-53), Aldershot, United Kingdom, Ashgate, 1997*, 11-28.

Klein, G. (1999). *Sources of power: How people make decisions*: MIT press.

Klein, G. (2004). *The power of intuition: How to use your gut feelings to make better decisions at work*: Crown Business.

Klein, G. (2008). Naturalistic decision making. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*(3), 456-460.

Klein, G., & Crandall, B. W. (1995). The role of mental simulation in naturalistic decision making. *Local applications of the ecological approach to human-machine systems, 2*, 324-358.

Klein, G., Orasanu, J., Calderwood, R., & Zsambok, C. (1993). *Decision making in action: Models and methods*.

Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2007). A data-frame theory of sensemaking. *Expertise out of context*, 113-155.

Klein, G., Shafer, J. L., & Ross, K. G. (2006). Professional Judgments and "Naturalistic Decision Making" *Cambridge Handbook of Expertise and Expert Performance*: Cambridge University Press.

Klein, G., & Wolf, S. (1998). The role of leverage points in option generation. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 28*(1), 157-160. doi: 10.1109/5326.661098

Klein, G., Wolf, S., Militello, L., & Zsambok, C. (1995). Characteristics of skilled option generation in chess. *Organizational Behavior and Human Decision Processes, 62*(1), 63-69.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. *Decision making in action: Models and methods, 5*(4), 138-147.

Klein, G. A., & Brezovic, C. P. (1986). *Design engineers and the design process: Decision strategies and human factors literature.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). *Rapid decision making on the fire ground.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *Systems, Man and Cybernetics, IEEE Transactions on, 19*(3), 462-472. doi: 10.1109/21.31053

Klein, G. A., & Hoffman, R. R. (1993). Perceptual-cognitive aspects of expertise. *Cognitive science foundations of instruction*, 203-226.

Kobayashi, H. (2005). Use of simulators in assessment, learning and teaching of mariners. *WMU Journal of Maritime Affairs, 4*(1), 57-75.

Kobus, D., Proctor, S., Bank, T., & Holste, S. (2000). Decision-making in a dynamic environment: The effects of experience and information uncertainty: DTIC Document.

Koester, T. (2003a). Psycho physiological measurements of mental activity, stress reactions and situation awareness in the maritime full mission simulator *Decision Making in Complex Environments* (pp. 311-320): Ashgate.

Koester, T. (2003b). Situation awareness and situation dependent behaviour adjustment in the maritime work domain. *Proceedings of HCI International, Crete, Greece*.

Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K.-R., Curio, G., . . . Kincses, W. (2007). Improving human performance in a real operating environment through real-time mental workload detection. *Toward Brain-Computer Interfacing*, 409-422.

Koschmann, T., LeBaron, C., Goodwin, C., & Feltovich, P. (2001). *Dissecting common ground: Examining an instance of reference repair.* Paper presented at the Proceedings of the Cognitive Science Society.

Koziol, L. F., Budding, D. E., & Chidekel, D. (2010). Adaptation, expertise, and giftedness: towards an understanding of cortical, subcortical, and cerebellar network contributions. *The Cerebellum, 9*(4), 499-529.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. *Multiple-task performance*, 279-328.

Kramer, A. F., Sirevaag, E. J., & Braune, R. (1987). A Psychophysiological Assessment of Operator Workload During Simulated Flight Missions. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 29*(2), 145-160. doi: 10.1177/001872088702900203

Krampe, R. T., & Ericsson, K. A. (1996). Maintaining excellence: Deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General, 125*(4), 331-359. doi: 10.1037/0882-7974.3.2.122326825010.1037/0882-7974.3.2.1221988-29162-001

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

Larjo, K. (2010). Practices in Pilotage – Past, Present and Future. *Multiprint Oy, Vantaa, Safety Study S1/2004M b*.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and Novice Performance in Solving Physics Problems. *Science, 208*(4450), 1335-1342. doi: 10.2307/1684057

Leavitt, J. (1979). Cognitive demands of skating and stickhandling in ice hockey. *Canadian journal of applied sport sciences. Journal canadien des sciences appliquées au sport, 4*(1), 46.

Lee, S., Black, A., Lacherez, P., & Wood, J. (2016). Eye Movements and Road Hazard Detection: Effects of Blur and Distractors. *Optometry and vision science: official publication of the American Academy of Optometry*.

Lei, S., Welke, S., & Roetting, M. (2009). Representation of driver's mental workload in EEG data. *Human Factors, security and safety*, 285-294.

Lesgold, A. M., & Resnick, L. B. (1982). How reading disabilities develop: Perspectives from a longitudinal study. *Theory and research in learning disability*, 155-187.

Lieberman, D. A. (2011). *Human learning and memory*: Cambridge University Press.

Lieberman, M. D., Jarcho, J. M., & Satpute, A. B. (2004). Evidence-based and intuition-based self-knowledge: An fMRI study. *Journal of Personality and Social Psychology, 87*(4), 421-435.

Lipshitz, R. (1993). Converging themes in the study of decision making in realistic settings. In J. O. G. A. Klein, R. Calderwood, & C. E. Zsambok (Eds.) (Ed.), *Decision making in action: Models and methods* (pp. 103-137). Westport, CT: Ablex Publishing.

Lipshitz, R. (2001). Puzzle-Seeking and Model Building on the Fire Ground: A Discussion of Karl Weick's Keynote Address'. *Lniking Expertise and Naturalistic Decision Making, Lawrence Erlbaum Associates, Publishers, Mahwah, NJ*, 321-336.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making, 14*(5), 331-352.

Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: A naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes, 69*(2), 149-163.

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 39*(2), 367-386. doi: 10.1080/00140137808931800 10.1037/0033-295x.84.2.127.

Lord, S., Senior, R., Das, M., Whittam, A., Murray, A., & McComb, J. (2001). Low-frequency heart rate variability: reproducibility in cardiac transplant recipients and normal subjects. *Clinical Science, 100*(1), 43-46.

Lützhöft, M., & Dukic, T. (2007). *Show me where you look and I'll tell you if you're safe: eye tracking of maritime watch-keepers.* Paper presented at the Proceedings of the 39th Nordic Ergonomics Society Conference.

Lützhöft, M. H., & Nyce, J. M. (2006). Piloting By Heart And By Chart. *The Journal of Navigation, 59*(02), 221-237. doi: doi:10.1017/S0373463306003663

Luximon, A., & Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. *Ergonomics, 44*(3), 229-243.

Lysaght, R. J., Hill, S. G., Dick, A., Plamondon, B. D., & Linton, P. M. (1989). Operator workload: Comprehensive review and evaluation of operator workload methodologies: DTIC Document.

Maddox, M., Wulf, G., & Wright, D. (1999). The effect of an internal vs. external focus of attention on the learning of a tennis stroke. *Journal of Sport and Exercise Psychology, 21*, 78.

Magee, L. (1997). Virtual Reality Simulator (VRS) for Training Ship Handling Skills. In R. Seidel & P. Chatelier (Eds.), *Virtual Reality, Training's Future?* (Vol. 6, pp. 19-29): Springer US.

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability standards of measurement, physiological interpretation, and clinical use. *European heart journal, 17*(3), 354-381.

Markman, A. B. (1999). *Knowledge Representation Pr*: Psychology Press.

MathWorks, I. (2013). MATLAB: the language of technical computing. Desktop tools and development environment: MathWorks.

McCarley, J. S., Wickens, C. D., Goh, J., & Horrey, W. J. (2002). *A computational model of attention/situation awareness.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). Job experience correlates of job performance. *Journal of Applied Psychology, 73*(2), 327-330.

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2009). A critical review of simulation based medical education research: 2003–2009. *Medical Education, 44*(1), 50-63.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica, 22*(3), 276-282.

McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology, 13*(3), 307-325. doi: 10.1016/0010-0285(81)90012-8

McKenna, F., & Crick, J. (1994). Hazard perception in drivers: A methodology for testing and training. *TRL Contractor Report*(313).

McKenna, F., & Crick, J. (1997). Developments in hazard perception. *TRL REPORT 297*.

McKenna, F., & Farrand, P. (1999). The role of automaticity in driving. In G. B. G. (Ed.) (Ed.), *Behavioural Research in Road Safety IX* (pp. 20-25). Crowthorne: Transport Research Laboratory.

McKinney, E. H., & Davis, K. J. (2003). Effects of Deliberate Practice on Crisis Decision Performance. *Human Factors, 45*(3), 436-444.

McNamara, R., Collins, A., & Mathews, V. (2000). A review of research into fatigue in offshore shipping. *Maritime review*, 118-122.

Meade, M. L., Nokes, T. J., & Morrow, D. G. (2009). Expertise promotes facilitation on a collaborative memory task. *Memory, 17*(1), 39-48.

Mieg, H. A. (2001). *The social psychology of expertise: Case studies in research, professional domains, and expert roles*: Lawrence Erlbaum.

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: a 15-year review of the team mental model construct. *Journal of Management, 36*(4), 876-910.
Monsaas, J. A. (1985). Learning to Be a World-Class Tennis Player. *Developing talent in young people*, 211.
Moray, N. (1986). Monitoring behavior and supervisory control. In K. R. Boff, Kaufman, L., & Thomas, J. P. (Eds.) (Ed.), *Handbook of perception and human performance* (Vol. Vol 2). New York: Wiley.
Moray, N. (1988). Mental workload since 1979. *International Reviews of Ergonomics, 2*, 123-150.
Mosier, K. (1991). *Expert decision-making strategies.* Paper presented at the International Symposium on Aviation Psychology, 6 th, Columbus, OH.
Moskowitz, G. B., Skurnik, I., & Galinsky, A. D. (1999). The history of dual-process notions, and the future of preconscious control. *Dual-process theories in social psychology*, 12-36.
Moulaert, V., Verwijnen, M. G., Rikers, R., & Scherpbier, A. J. (2004). The effects of deliberate practice in undergraduate medical education. *Medical Education, 38*(10), 1044-1052.
Mourant, R. R., & Rockwell, T. H. (1972). Strategies of visual search by novice and experienced drivers. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 14*(4), 325-335.
Muczyński, B., Gucma, M., Bilewski, M., & Zalewski, P. (2013). Using eye tracking data for evaluation and improvement of training process on ship's navigational bridge simulator. *Zeszyty Naukowe/Akademia Morska w Szczecinie*(33 (105)), 75--78.
Mulder, G. (1980). *The heart of mental effort: Studies in the cardiovascular psychophysiology of mental work*. Groningen.
Mulder, L. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology, 34*(2), 205-236.
Mulgund, S. S., Harper, K. A., Zacharias, G. L., & Menke, T. (2000). *SAMPLE: situation awareness model for pilot-in-the-loop evaluation.* Paper presented at the 9th Conference on Computer Generated Forces and Behavioral Representation.
Mumford, M. D., Hester, K. S., Robledo, I. C., Peterson, D. R., Day, E. A., Hougen, D. F., & Barrett, J. D. (2012). Mental models and creative problem-solving: The relationship of objective and subjective model attributes. *Creativity Research Journal, 24*(4), 311-330.
Murai, K., Hayashi, Y., Nagata, N., & Inokuchi, S. (2004). The mental workload of a ship's navigator using heart rate variability. *Interactive Technology and Smart Education, 1*(2), 127-133.
Nagle, F., Naughton, J., & Balke, B. (1966). Comparisons of direct and indirect blood pressure with pressure-flow dynamics during exercise. *Journal of Applied Physiology, 21*(1), 317-320.
Nasim, K., Jahan Ara, H., & Syed Sanowar, A. (2011). Heart rate variability  - a review. *Journal of Basic & Applied Sciences, 7*(1), n/a-n/a.
Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86*(3), 214.
Naweed, A., & Balakrishnan, G. (2012). Perceptions and experiences of simulators as a training tool in transport: The case of the Australian rail industry. *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice, 21*(3), 77.
Neisser, U. (1976). *Cognition and Reality: Principles and Implications of Cognitive Psychology*. New York: WH Freeman and Co.
Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Commun. ACM, 19*(3), 113-126. doi: 10.1145/360018.360022

Nickel, P., & Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 45*(4), 575-590.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychological Review, 108*(2), 291.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*(1), 1-32.

Noyes, J., Garland, K., & Robbins, L. (2004). Paper-based versus computer-based assessment: is workload another test mode effect? *British Journal of Educational Technology, 35*(1), 111-113. doi: 10.1111/j.1467-8535.2004.00373.x

O'Hare, D. (1997). Cognitive ability determinants of elite pilot performance. *Human Factors, 39*(4), 540-540.

Oermann, M. H., Kardong-Edgren, S., Odom-Maryon, T., Hallmark, B. F., Hurd, D., Rogers, N., . . . Smart, D. A. (2011). Deliberate practice of motor skills in nursing education: cpr as exemplar. *Nursing Education Perspectives, 32*(5), 311-315.

Orlandi, L., & Brooks, B. (2018). Measuring mental workload and physiological reactions in marine pilots: Building bridges towards redlines of performance. *Applied Ergonomics, 69*, 74-92. doi: https://doi.org/10.1016/j.apergo.2018.01.005

Orlandi, L., Brooks, B., & Bowles, M. (2014). *The development of a shiphandling assessment tool (SAT): A methodology and an integrated approach to assess manoeuvring expertise in a full mission bridge simulator.* Paper presented at the 15th Annual general assembly International Association of Maritime Universities.

Orlandi, L., Brooks, B., & Bowles, M. (2015). A Comparison of Marine Pilots' Planning and Manoeuvring Skills: Uncovering Mental Models to Assess Shiphandling and Explore Expertise. *Journal of Navigation, 68*(5), 897-914.

Orlandi, L., Brooks, B., Wood, J., & Black, A. (201x). Interpreting changes in marine pilots' perceptual cycle through gaze detection and speech patterns. *Ergonomics, Under Review*.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140-160.

Pascual, R., & Henderson, S. (1997). Evidence of naturalistic decision making in military command and control. *Naturalistic decision making(A 97-39376 10-53), Mahwah, NJ, Lawrence Erlbaum Associates, Inc., Publishers, 1997*, 217-226.

Patel, V. L., Arocha, J. F., & Kaufman, D. R. (1994). Diagnostic reasoning and medical expertise. *The psychology of learning and motivation, 31*, 187-252.

Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. *Toward a general theory of expertise*, 93-125.

Patil, R. S., Szolovits, P., & Schwartz, W. B. (1981). *Causal understanding of patient illness in medical diagnosis.* Paper presented at the Proceedings of the Seventh International Joint Conference on Artificial Intelligence.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 534.

Pelz, D. C., & Krupat, E. (1974). Caution profile and driving record of undergraduate males. *Accident Analysis & Prevention, 6*(1), 45-58. doi: 10.1016/0001-4575(74)90015-3

Penaz, J. (1978). Mayer waves: history and methodology. *Automedica, 2*, 135-141.

Perkovic, M., Brcko, T., Luin, B., & Vidmar, P. (2016). *Ship handling challenges when vessels are outgrowing ports.* Paper presented at the 19th International Navigation Simulator Lecturers' Conference–INSLC.

Perrow, C. (2011). *Normal accidents: Living with high risk technologies*: Princeton University Press.

Pham, M. T. (2004). The logic of feeling. *Journal of Consumer Psychology, 14*(4), 360-369.

Phelps, R. H., & Shanteau, J. (1978). Livestock judges: How much information can an expert use? *Organizational Behavior and Human Performance, 21*(2), 209-219. doi: 10.1016/0030-5073(78)90050-8

Phillips, J. K., Klein, G., & Sieck, W. R. (2004). Expertise in judgment and decision making: A case for training intuitive decision skills. *Blackwell Handbook of Judgement and Decision-making, Blackwell, Oxford*, 297-315.

PIANC. (2014). Standardization of ships and inland waterways for river/sea navigation. *MarCom Working Group 12, MarCom report 121*(January 2014).

Plant, E. A., Ericsson, K. A., Hill, L., & Asberg, K. (2005). Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance. *Contemporary Educational Psychology, 30*(1), 96-116. doi: 10.1016/j.cedpsych.2004.06.001

Plant, K. L., & Stanton, N. A. (2012). Why did the pilots shut down the wrong engine? Explaining errors in context using Schema Theory and the Perceptual Cycle Model. *Safety Science, 50*(2), 300-315.

Plant, K. L., & Stanton, N. A. (2013). What is on your mind? Using the perceptual cycle model and critical decision method to understand the decision-making process in the cockpit. *Ergonomics, 56*(8), 1232-1250.

Plimpton, G. (1977). *The Paris Review Interviews: Writers at Work, 5 vols*. New York: Penguin Books.

Posner, M. I., & Snyder, C. R. R. (2004). Attention and Cognitive Control. *Cognitive psychology: Key readings*, 205.

Postal, V. (2004). Expertise in Cognitive Psychology: Testing the Hypothesis of Long-term Working Memory in a Study of Soccer Players. *Perceptual and Motor Skills, 99*(2), 403-420.

Press, M. (1986). Situation awareness: Let's get serious about the clue-bird. *Unpublished manuscript*.

Quimby, A. R. (1987). Perceptual abilities of accident-involved drivers. *Journal of Safety Research, 18*(1), 45. doi: 10.1016/0022-4375(87)90068-5

Ramshur, J. T. (2010). HRVAS: Heart Rate Variability Analysis Software. *University of Memphis. Department of Biomedical Engineering*.

Randel, J. M., Pugh, H. L., & Reed, S. K. (1996). Differences in expert and novice situation awareness in naturalistic decision making. *International Journal of Human-Computer Studies, 45*(5), 579-597. doi: 10.1006/ijhc.1996.0068

Raskin, E. (1925). Comparison of scientific and literary ability: a biographical study of eminent scientists and men of letters of the nineteenth century. *The Journal of Abnormal and Social Psychology, 31*(1), 20-35.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE transactions on systems, man, and cybernetics*(3), 257-266.

Reid, G. B., Shingledecker, C. A., Nygren, T. E., & Eggemeier, F. T. (1981). *Development of multidimensional subjective measures of workload.* Paper presented at the Proceedings of the IEEE International Conference on Cybernetics and Society.

Reif, F., & Allen, S. (1992). Cognition for Interpreting Scientific Concepts: A Study of Acceleration. *Cognition and Instruction, 9*(1), 1-44. doi: 10.2307/3233621

Reitman, J. S. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology, 8*(3), 336-356. doi: 10.1016/0010-0285(76)90011-6

Rieger, M. (2004). Automatic Keypress Activation in Skilled Typing. [Article]. *Journal of Experimental Psychology. Human Perception & Performance, 30*(3), 555-565. doi: 10.1037/0096-1523.30.3.555

Riley, M. A., Stoffregen, T. A., Grocki, M. J., & Turvey, M. (1999). Postural stabilization for the control of touching. *Human Movement Science, 18*(6), 795-817.

Ritvanen, T., Louhevaara, V., Helin, P., Väisänen, S., & Hänninen, O. (2006). Responses of the autonomic nervous system during periods of perceived high and low work stress in younger and older female teachers. *Applied Ergonomics, 37*(3), 311-318.

Roberts, K. H. (1990). Some Characteristics of One Type of High Reliability Organization. *Organization Science, 1*(2), 160-176. doi: 10.1287/orsc.1.2.160

Rompelman, O., Coenen, A. J., & Kitney, R. (1977). Measurement of heart-rate variability: Part 1—Comparative study of heart-rate variability analysis methods. *Medical and Biological Engineering and Computing, 15*(3), 233-239.

Rosch, J. L., & Vogel-Walcutt, J. J. (2013). A review of eye-tracking applications as tools for training. *Cognition, technology & work, 15*(3), 313-327.

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology, 34*(2–3), 259-287. doi: 10.1016/0301-0511(92)90018-p

Rosenbaum, D. A., Augustyn, J. S., Cohen, R. G., & Jax, S. A. (2006). Perceptual-Motor Expertise *Cambridge Handbook of Expertise and Expert Performance*: Cambridge University Press.

Ross, K., McHugh, A., Moon, B., Klein, G., Armstrong, A., & Rall, E. (2002). High-level cognitive processes in field research (Year One Final Report under Contract 02TA2-SP1-RT1 for US Army Research Laboratory under Cooperative Agreement DAAD19–01–2–0009). *Fairborn, OH: Klein Associates*.

Ross, K. G., Battaglia, D., Phillips, J., Domeshek, E. A., & Lussier, J. W. (2003). *Mental models underlying tactical thinking skills.* Paper presented at the The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

Rosson, M. B. (1985). *The role of experience in editing.* Paper presented at the Interact.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100*(3), 349-363.

Rutherford, A., & Wilson, J. R. (1989). *Models of mental models: an ergonomist-psychologist dialogue.* Paper presented at the Selected papers of the 8th Interdisciplinary Workshop on Informatics and Psychology: Mental Models and Human-Computer Interaction 2.

Ryan, N. K. (1998). *The future of maritime facility designs and operations.* Paper presented at the Simulation Conference Proceedings, 1998. Winter.

Rydstedt, L. W., & Lundh, M. (2010). An ocean of stress? The relationship between psychosocial workload and mental strain among engine officers in the Swedish merchant fleet. *International maritime health, 62*(3), 168-175.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It Is Not How Much You Have but How You Use It: Toward a Rational Use of Simulation to Support Aviation Training. *The International Journal of Aviation Psychology, 8*(3), 197-208. doi: 10.1207/s15327108ijap0803_2

Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors, 37*(1), 123-136.

Salas, E., Rosen, M. A., & DiazGranados, D. (2010). Expertise-based intuition and decision making in organizations. *Journal of Management, 36*(4), 941-973.

Salmon, P. M., Stanton, N. A., Walker, G. H., Baber, C., Jenkins, D. P., McMaster, R., & Young, M. S. (2008). What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science, 9*(4), 297-323.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics, 39*(3), 490-500.

Salthouse, T. A. (1984). Effects of age and skill in typing. *Journal of Experimental Psychology: General, 113*(3), 345-371. doi: 10.1016/s0166-4115(08)61962-710.1016/s0166-4115(08)61962-7 10.1080/00140137808931800

Salthouse, T. A. (1991). Expertise as the circumvention of human processing limitations. *Toward a general theory of expertise: Prospects and limits*, 286.

Sampson, H. (2004). Romantic rhetoric, revisionist reality: the effectiveness of regulation in maritime education and training. *Journal of Vocational Education and Training, 56*(2), 245-267.

Sampson, H., & Bloor, M. (2007). When Jack Gets out of the Box: The Problems of Regulating a Global Industry. *Sociology, 41*(3), 551-569. doi: 10.1177/0038038507076623

Sampson, H., Bloor, M., & Gekara, V. (2011). Water-tight or sinking? A consideration of the standards of the contemporary assessment practices underpinning seafarer licence examinations and their implications for employers. [Article]. *Maritime Policy and Management, 38*(1), 81-92. doi: 10.1080/03088839.2010.533713

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 49*(3), 347-357.

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 5-19.

Savage, I. (2013). Comparing the fatality risks in United States transportation across modes and over time. *Research in transportation economics, 43*(1), 9-22.

Scardamalia, M., & Bereiter, C. (1991). Literate expertise. *Toward a general theory of expertise: Prospects and limits*, 172-194.

Schank, R. C., & Colby, K. M. (1973). *Computer models of thought and language*: WH Freeman & Co.

Schmidt, J. A. (1989). Novice Strategies for Understanding Paintings. *Applied Cognitive Psychology, 3*(1), 65-72. doi: 10.1002/acp.2350030107

Schmitt, J. F., & Klein, G. (1999). How we plan. *Marine Corps Gazette, 83*, 18-26.

Schmitt, J. F., & Klein, G. A. (1998). Fighting in the fog: Dealing with battlefield uncertainty. *Human Performance in Extreme Environments; Human Performance in Extreme Environments*.

Schneider, W., & Fisk, A. D. (1982). Concurrent automatic and controlled visual search: Can processing occur without resource cost? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(4), 261-278. doi: 10.1037/h0043158.1957-02914-00110.1037/h0043158 10.1037/0033-295x.84.2.127.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1-66.

Schulz, R., & Curnow, C. (1988). Peak performance and age among superathletes: Track and field, swimming, baseball, tennis, and golf. *Journal of gerontology, 43*(5), P113-P120.

Selye, H. (1980). *Selye's guide to stress research*: Van Nostrand Reinhold Company.

Shaffer, L. (1975). Multiple attention in continuous verbal tasks. *Attention and performance V*, 157-167.

Shanteau, J. (1989). Cognitive heuristics and biases in behavioral auditing: Review, comments and observations. *Accounting, Organizations and Society, 14*(1), 165-177.

Shanteau, J. (1992). The Psychology of Experts An Alternative View. *Expertise and decision support*, 11-23.
Shanteau, J., & Stewart, T. R. (1992). Why study expert decision making? Some historical perspectives and comments. *Organizational Behavior and Human Decision Processes, 53*(2), 95-106. doi: 10.1016/0749-5978(92)90057-E
Shattuck, L., Graham, J., Merlo, J., & Hah, S. (2000). *Cognitive Integration: An investigation of how expert and novice commanders process battlefield data.* Paper presented at the Fourth Annual Federated Laboratory Symposium on Advanced Displays and Interactive Displays Consortium, Adelphi, MD.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127-190.
Shively, R. J., Brickner, M., & Silbiger, J. (1997). *A computational model of situational awareness instantiated in MIDAS(Man-machine Integration Design and Analysis).* Paper presented at the International Symposium on Aviation Psychology, 9 th, Columbus, OH.
Sidney Dekker, E. H., David Woods, & Cook, a. R. (2008). Resilience Engineering: New directions for measuring and maintaining safety in complex systems: Lund University School of Aviation.
Simon, H. A. (1957). *Models of man; social and rational*. Oxford, England: Wiley.
Simon, H. A. (1989). *Models of Thought, volume II*. New Haven, CT: Yale University Press.
Simon, H. A., & Chase, W. G. (1973). Skill in Chess: Experiments with chess-playing tasks and computer simulation of skilled performance throw light on some human perceptual and memory processes. *American scientist, 61*(4), 394-403. doi: 10.2307/27843878
Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology, 16*(1), 97-115. doi: 10.1002/acp.759
Singer, R. N. (2002). Preperformance state, routines and automaticity: What does it take to realize expertise in self-paced events? *Journal of Sport & Exercise Psychology, 24*(4), 359-375.
Sirevaag, E. J., Kramer, A. F., Reisweber, C. D. W. M., Strayer, D. L., & Grenell, J. F. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics, 36*(9), 1121-1140.
Skinner, M. J., & Simpson, P. A. (2002). Workload Issues in Military Tactical Airlift. *The International Journal of Aviation Psychology, 12*(1), 79-93. doi: 10.1207/s15327108ijap1201_7
Sloboda, J. A. (1996). The role of practice in the development of performing musicians. *The British journal of psychology, 87*(2), 287-309. doi: 10.1111/j.2044-8295.1996.tb02591.x
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.
Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral Decision Theory. *Annual Review of Psychology, 28*(1), 1-39. doi: doi:10.1146/annurev.ps.28.020177.000245
Smith, J. F., & Kida, T. (1991). Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin, 109*(3), 472-489. doi: 10.1016/0022-1031(78)90021-5
Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 137-148.
Smith, M. E., Gevins, A., Brown, H., Karnik, A., & Du, R. (2001). Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. *Human Factors, 43*(3), 366-380.

Smith, P., Shah, M., & da Vitoria Lobo, N. (2000). *Monitoring head/eye motion for driver alertness with one camera.* Paper presented at the Pattern Recognition, 2000. Proceedings. 15th International Conference on.

Sonnentag, S. (1998). Expertise in professional software design: A process study. *Journal of Applied Psychology, 83*(5), 703-715. doi: 10.1016/s0364-0213(87)80026-5.1988-19111-00110.1016/s0364-0213(87)80026-5

Sonnentag, S., Niessen, C., & Volmer, J. (2009). Expertise in software design. In K. A. C. Ericsson, Neil; Feltovich, Paul J.; Hoffmann, Robert R. (ed.) (Ed.), *Cambridge handbook of expertise and expert performance* (pp. 373-387). Cambridge: Cambridge University Press.

Sosniak, L. A. (1985). Learning to be a concert pianist. *Developing talent in young people*, 19-67.

Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition, 4*(3), 215-230. doi: 10.1016/0010-0277(76)90018-4

Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*(3), 275-290. doi: 10.1016/S0022-5371(79)90155-5

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*: Psychology Press.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*(5), 645-665.

Stanton, N., Salmon, P. M., & Rafferty, L. A. (2013). *Human factors methods: a practical guide for engineering and design*: Ashgate Publishing, Ltd.

Stanton, N. A. (2016). Distributed situation awareness. *Theoretical issues in ergonomics science*. doi: 10.1080/1463922X.2015.1106615

Stanton, N. A., Salmon, P. M., Walker, G. H., & Jenkins, D. P. (2010). Is situation awareness all in the mind? *Theoretical Issues in Ergonomics Science, 11*(1-2), 29-40.

Starcke, K., & Brand, M. (2012). Decision making under stress: A selective review. *Neuroscience & Biobehavioral Reviews, 36*(4), 1228-1248. doi: 10.1016/j.neubiorev.2012.02.003

Starkes, J. L., Deakin, J. M., Allard, F., & Hodges, N. J. (1996). Deliberate practice in sports: What is it anyway? In K. A. Ericsson (Ed.), *The Road To Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games* (pp. 81). New York: Psychology Press.

Staszewski, J. I., & Simon, H. A. (1995). Perceptual and Memory Processes in the Acquisition of Expert Performance: The EPAM Model. In K. A. Ericsson (Ed.), *The Road To Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games* (pp. 167-187). New York: Psychology Press.

STCW, I. (2011). International Convention on Standards of Training, Certification and Watchkeeping for Seafarers,(STCW) 1978, as amended in 1995/2010. *International Maritime Organisation, London, UK*.

Sterman, M. B., Mann, C. A., Kaiser, D. A., & Suyenobu, B. Y. (1994). Multiband topographic EEG analysis of a simulated visuomotor aviation task. *International Journal of Psychophysiology, 16*(1), 49-56.

Stevens, R. H., Galloway, T., & Berka, C. (2007). *EEG-related changes in cognitive workload, engagement and distraction as students acquire problem solving skills.* Paper presented at the International Conference on User Modeling.

Stone, B., Lee, M., Dennis, S., & Nettelbeck, T. (2004). *Pupil size and mental load.* Paper presented at the 1st Adelaide Mental Life Conference, Adelaide.

Stout, R. J., Cannon-Bowers, J. A., Salas, E., & Milanovich, D. M. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established.

*Human Factors: The Journal of the Human Factors and Ergonomics Society, 41*(1), 61-71.

Strater, L. D., Endsley, M. R., Pleban, R. J., & Matthews, M. D. (2001). Measures of platoon leader situation awareness in virtual decision-making exercises: DTIC Document.

Strater, L. D., Jones, D., & Endsley, M. R. (2001). Analysis of infantry situation awareness training requirements: DTIC Document.

Sun, H., Zimmer, H. D., & Fu, X. (2011). The influence of expertise and of physical complexity on visual short-term memory consolidation. *The Quarterly journal of experimental psychology, 64*(4), 707-729.

Swets, J. A. (2014). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*: Psychology Press.

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-Aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV–heart rate variability analysis software. *Computer Methods and Programs in Biomedicine, 113*(1), 210-220.

Thackray, R. I. (1969). *Patterns of physiological activity accompanying performance on a perceptual-motor task*: Department of Transportation, Federal Aviation Administration, Office of Aviation Medicine.

Thayer, J. F., Åhs, F., Fredrikson, M., Sollers Iii, J. J., & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews, 36*(2), 747-756. doi: 10.1016/j.neubiorev.2011.11.009

Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G.-Z., & Darzi, A. (2014). Eye tracking for skills assessment and training: a systematic review. *journal of surgical research, 191*(1), 169-178.

Uhlarik, J., & Comerford, D. A. (2002). *A review of situation awareness literature relevant to pilot surveillance functions*: DIANE Publishing.

Underwood, G., Chapman, P., Bowden, K., & Crundall, D. (2002). Visual search while driving: skill and awareness during inspection of the scene. *Transportation Research Part F: Traffic Psychology and Behaviour, 5*(2), 87-97.

USArmy. (1997). *Field Manual 101-5 - Staff Organization and Operations*.

USMarineCorps. (1998). Marine Corps Planning Process. *US Marine Corps*.

Valentin, D., Pichon, M., de Boishebert, V., & Abdi, H. (2000). What's in a wine name? When and why do wine experts perform better than novices. *Psychonomic Society, 5*, 36.

Van Amelsvoort, L., Schouten, E., Maan, A., Swenne, C., & Kok, F. (2000). Occupational determinants of heart rate variability. *International Archives of Occupational and Environmental Health, 73*(4), 255-262.

Van Steenis, H., Tulen, J., & Mulder, L. (1994). Heart rate variability spectra based on non-equidistant sampling: the spectrum of counts and the instantaneous heart rate spectrum. *Medical engineering & physics, 16*(5), 355-362.

Van Westrenen, F. C. (1999). *The maritime pilot at work: Evaluation and use of a time-to-boundary model of mental workload in human-machine systems.* (Dr. C719584), Technische Universiteit Delft (The Netherlands), Netherlands. ProQuest Dissertations & Theses A&I database.

Veritas, D. N. (2011). Standard for Certification No. 2.14 Maritime Simulator Systems. *Det Norske Veritas (DNV) Standards for Certification*.

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour, 8*(2), 167-190.

Vidulich, M., Dominguez, C., Vogel, E., & McMillan, G. (1994). Situation awareness: Papers and annotated bibliography: DTIC Document.

Vlek, C. (1984). What constitutes ‘a good decision’?: A panel discussion among Ward Edwards, István Kiss, Giandomenico Majone and Masanao Toda. *Acta Psychologica, 56*(1–3), 5-27. doi: 10.1016/0001-6918(84)90004-0
Von Neumann, J., & Morgenstern, O. (2007). *Theory of games and economic behavior (commemorative edition)*: Princeton university press.
Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem-solving skill in the social sciences. *The psychology of learning and motivation, 17*, 165-213.
Wallingford, R. (1975). Long distance running. *The scientific aspects of sport training. Springfield: Charles C. Thomas*, 118-130.
Waterman, D. (1986). *A guide to expert systems*. Reading, MA: Addison-Wesle.
Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72*(2), 89-104.
Weber, N., & Brewer, N. (2003). Expert memory: The interaction of stimulus structure, attention, and expertise. *Applied Cognitive Psychology, 17*(3), 295-308.
Weibel, N., Fouse, A., Emmenegger, C., Kimmich, S., & Hutchins, E. (2012). *Let's look at the cockpit: exploring mobile eye-tracking for observational research on the flight deck.* Paper presented at the Proceedings of the Symposium on Eye Tracking Research and Applications.
Weick, K. E. (1993). Sensemaking in organizations: Small structures with large consequences. *Social psychology in organizations: Advances in theory and research*, 10-37.
Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3): SAGE Publications, Incorporated.
Weick, K. E., & Roberts, K. H. (1993). Collective mind in organizations: Heedful interrelating on flight decks. *Administrative science quarterly*, 357-381.
Weiss, S. M., & Kulikowski, C. A. (1984). *A practical guide to designing expert systems*: Rowman & Littlefield.
Welch, P. D. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics, 15*(2), 70-73.
Welford, A. T. (1968). *Fundamentals of skill*. New York: Methuen.
Wickens, C. (1992). Workload and situation awareness: An analogy of history and implications. *Insight: The visual performance technical group newsletter, 14*(4), 1-3.
Wickens, C. D. (1984). *Engineering psychology and human performance*: Merrill.
Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159-177.
Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*(3), 449-455.
Wickens, C. D., Gordon, S. E., Liu, Y., & Lee, J. (1998). *An introduction to human factors engineering*. Upper Saddle River, New Jersey: Pearson Education, Inc.
Wieland-Eckelmann, R. (1992). *A composite model of multiple-task performance: Theory and empirical data*: Fach Psychologie, Fachber. 3, Univ.-GH.
Wierwille, W. W. (1979). Physiological measures of aircrew mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 21*(5), 575-593.
Wierwille, W. W. (1988). Important remaining issues in mental workload estimation. *Advances in Psychology, 52*, 315-327.
Wild, R., & Constable, K. (2013). A Document Of Debatable Value – A Case Study Into The Use Of Master-Pilot Exchange Documentation In Selected UK Ports. *The Journal of Navigation, 66*(03), 465-471. doi: doi:10.1017/S0373463313000040
Williams, A. M., Ward, P., Knowles, J. M., & Smeeton, N. J. (2002). Anticipation skill in a real-world task: Measurement, training, and transfer in tennis. *Journal of Experimental Psychology: Applied, 8*(4), 259-270. doi: 10.1037/0033-2909.124.3.372
Williams, M. (2004). *Skill acquisition in sport: Research, theory and practice*: Routledge.

Williams, M., & Davids, K. (1995). Declarative knowledge in sport: a by-product of experience or a characteristic of expertise? *Journal of Sport & Exercise Psychology, 17*(3), 259-275.

Wilson, G. F. (2002). An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology, 12*(1), 3-18. doi: 10.1207/s15327108ijap1201_2

Wilson, J. R., & Rutherford, A. (1989). Mental models: Theory and application in human factors. *Human Factors, 31*(6), 617-634.

Woods, D. D., & Sarter, N. B. (2010). Capturing the dynamics of attention control from individual to distributed systems: the shape of models to come. *Theoretical Issues in Ergonomics Science, 11*(1-2), 7-28.

Woodworth, R., & Schlosberg, H. (1954). *Experimental psychology*: Holt.

Wu, Y., Miwa, T., & Uchida, M. (2017). Using physiological signals to measure operator's mental workload in shipping–an engine room simulator study. *Journal of Marine Engineering & Technology, 16*(2), 61-69.

Wulf, G., Lauterbach, B., & Toole, T. (1999). The learning advantages of an external focus of attention in golf. *Research Quarterly for Exercise and Sport, 70*, 120-126.

Wulf, G., & Prinz, W. (2001). Directing attention to movement effects enhances learning: A review. *Psychonomic Bulletin & Review, 8*(4), 648-660.

Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S., & Cook, K. S. (2009). The private rejection of unfair offers and emotional commitment. *Proceedings of the National Academy of Sciences, 106*(28), 11520-11523. doi: 10.1073/pnas.0900636106

Yamamoto, S., & Matsuoka, S. (1990). Topographic EEG study of visual display terminal (VDT) performance with special reference to frontal midline theta waves. *Brain topography, 2*(4), 257-267.

Yates, F. J., & Tschirhart, M. D. (2006). Decision-Making Expertise *Cambridge Handbook of Expertise and Expert Performance*: Cambridge University Press.

Yates, J. F. (1990). *Judgment and decision making*: Prentice-Hall, Inc.

Yates, J. F., Veinott, E., & Patalano, A. L. (2003). Hard decisions, bad decisions: On decision quality and decision aiding. *Emerging perspectives on judgment and decision research*, 13-63.

Yemao, M., Lundh, M., & Porathe, T. (2014). Seeking harmony in shore-based unmanned ship handling-from the perspective of human factors, what is the difference we need to focus on from being onboard to onshore. *Advances in Human Aspects of Transportation. Part I, 231239*.

Young, M., & Stanton, N. (2005). Mental workload. *Handbook of human factors and ergonomics methods*, 39-31.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics, 58*(1), 1-17.

Yurko, Y. Y., Scerbo, M. W., Prabhu, A. S., Acker, C. E., & Stefanidis, D. (2010). Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simulation in healthcare, 5*(5), 267-271.

Zacharias, G. L., Miao, A. X., Illgen, C., Yara, J. M., & Siouris, G. M. (1996). *SAMPLE: Situation awareness model for pilot-in-the-loop evaluation.* Paper presented at the First Annual Conference on Situation Awareness in the Tactical Air Environment.

Zeitz, C. M. (1997). *Some concrete advantages of abstraction: How experts' representations facilitate reasoning.* Paper presented at the Expertise in context.

Zijlstra, F., & Van Doorn, L. (1985). *The construction of a scale to measure perceived effort*: University of Technology.

Zijlstra, F. R. H. (1993). *Efficiency in work behaviour: A design approach for modern tools (Ph. D. Thesis).* Delft University. The Netherlands, Delft University Press. Retrieved from https://repository.tudelft.nl/islandora/object/uuid:d97a028b-c3dc-4930-b2ab-a7877993a17f/datastream/OBJ

Zsambok, C., & Klein, G. (2014). Naturalistic decision making: where are we now *Naturalistic Decision Making.* Hillsdale, NJ: Psychology Press.

# APPENDICES

## APPENDIX 1: ADVERTISEMENT LETTER

AUSTRALIAN MARITIME COLLEGE

**SEA Seafarer Expertise Assessment – Expertise in Maritime Pilotage**
**PhD Research**

Dear Captain,

I would like to have the honour and the pleasure to invite you to participate in a study about Expertise in the Maritime Domain.
The research is being coordinated by the Australian Maritime College (AMC) with the involvement of the Brisbane Marine Pilots (BMP), the Australasian Marine Pilots Institute (AMPI) and the Marine Safety Queensland Smartship Simulator. The research will be conducted in fulfilment of a PhD course that I am attending, under the supervision of Dr. Ben Brooks (AMC).

**What is the purpose of this study?**
The aim of this study is to deepen the actual knowledge about Pilot Expertise.
The study, using an integrated and unobtrusive approach, will take into account performance, decision making and situational awareness, while Pilots are engaged in conducting ship operations. For this purpose a methodology will be employed, that will exploit different measurements, such as technical indicators and physiological parameters.

**Why have you been invited to participate in this study?'**
Brisbane Marine Pilots have always been keen to explore and integrate the most recent technologies and techniques in their approach to the profession.
You are the most welcome to participate in this study as an active member of the Brisbane Marine Pilots.

Offering your participation, you could actively contribute to the achievement of safer maritime operations, helping to deepen the knowledge of more efficient and effective practices in Pilotage.

**What does this study involve?**
Professional skills and techniques shown by Pilots will be observed during simulated and real manoeuvres on board of ships. Data collection will consider performance outcomes with regard to the safety, accuracy and efficacy of ship conduction. It will be taken into account how participants will deal with: the use of Bridge applications and equipment, such as Electronic Chart Display and Information System (ECDIS), Integrated Navigation Systems (INS), radars, electronic and physical navigation aids, consideration of natural and weather conditions present at the time, static and dynamic characteristics of the vessel influencing its manoeuvrability and evaluation of the environmental and infrastructural constraints of the Port...
A range of physiological data will be collected during the experiments, both in the simulated environment and on board ships. These data collection will include physiological measures like Eye tracking, Electroencephalography, Heart Rate Variability, Respiration Variability.

**Are there any possible benefits from participation in this study?**

An important element of this research will be the exploration of the qualitative and quantitative use made by the Pilots of all the previously mentioned systems, aids and information, in order to provide successive criteria to identify possible improvements to human machine interfaces, communications and operational procedures at different levels and standards.

The research could help answering at important questions such as:

What is the level of safety and resiliency granted nowadays in port operations, through the integration of complex systems such as Pilot – Ships – Tugs – VTS (Port Control) – Port logistics and services?

What is the level of expertise required and what are the criteria and the parameters to be considered in order to maintain proper and reliable safety standards in Port operations?

Your participation in this study will certainly enhance your awareness of these issues and it will ultimately contribute to a safer and more efficient working environment.

**Are there any possible risks from participation in this study?**

There are no foreseen risks as the required tasks will be carried out in a simulated environment or during routinely operations.

It is important to say that the **study results will remain completely anonymous**.

Any label or individual identifier will be permanently removed, so that, by no mean, a specific individual can be identified. Stored data will be able to be linked with other data so it can be known that they are about the same subject, but the person's identity will remain completely unknown.

Audio, video and physiological data recordings, are functional to the measurement and analysis. Data collected at Smartship and on-board ships will be shared with the AMC researchers for further analysis throughout the PhD programme. All findings will be securely stored in The Australian Maritime College at the University of Tasmania for 5 years. The computer files will be kept for comparison with future studies of crew performance. No video or audio recording will be published unless directly and specifically authorized, or deliberately masked in order to deceive individual identity.

During your participation, you may discontinue at any time, without providing any explanation.

**What if you have questions about this research?**

If you would like to discuss any aspect of this study, you are the most welcome to contact me at the contact details provided below.

If you wish to take part in the research, please feel free to confirm your interest, contacting me at the same accounts.

Thank you for taking the time to read this letter and consider this study.

Looking forward to having the pleasure to share with you this amazing experience, I take the opportunity to send my

Best Regards

Luca Orlandi

National Centre for Ports and Shipping
Australian Maritime College
University of Tasmania

E: lucaorlandi74@gmail.com
luca.orlandi@utas.edu.au

# PARTICIPANT INFORMATION SHEET

## SEA Seafarer Expertise Assessment – Expertise in Maritime Pilotage PhD research

**Invitation**

You are invited to participate in a study of Expertise Assessment in the maritime domain. The research is being coordinated by the Australian Maritime College (AMC) with involvement from the Brisbane Marine Pilots (BMP) and the Marine Safety Queensland Smartship Simulator. The study is being conducted in partial fulfilment of a PhD for Luca Orlandi under the supervision of Dr. Ben Brooks (AMC).

- **What is the purpose of this study?**

The area of research of this study is about the general knowledge regarding how to evaluate and assess, using an integrated and unobtrusive approach, Pilot's performance, decision making and situational awareness, while engaged in conducting ship operations.
For this purpose an assessing methodology will be employed, that will take into account correlations of different physiological, behavioral markers with pilot's performance.
The assessment of seafarers performance will include technical and non technical skills identified by a Behavioural Markers System and Physiological Markers System.

- **'Why have I been invited to participate in this study?'**

You are invited to participate in this study as an active member of the Brisbane Marine Pilots. Your participation will be voluntary and if you will decide to participate, you will have the possibility to withdraw at any time and at any stage of the study.

- **'What does this study involve?'**

The research will be divided into two phases. The first phase will be conducted in the Simulator, using two simulated scenarios: Brisbane Port and the Imaginary Port of "Vorbasse", a simulated non existent Port. The second phase, will consist of observations carried out on the field, during real mooring operations.
The first phase will include a set of four experiments carried out in the two mentioned scenarios. Each experiment will consist in a ship's movement or "manoeuvre". Within each scenario, the manoeuvres will be differentiated by levels of complexity. To give an idea, the simplest experiment will consist in a mooring with a small vessel without current or wind, the most difficult will include the presence of current and wind, a bigger vessel, the use of tugs, a possible failure in the equipment. All the exercises represent possible scenarios that may happen to a Pilot in real port operations.
Data collection will consider the human element and performance with regard to the safety, accuracy and efficacy of ship conduction. It will be taken into account how the Pilots will deal with: the use of Bridge applications and equipment, such as Electronic Chart Display and Information System (ECDIS), Integrated Navigation Systems (INS), radars, electronic and physical navigation aids, consideration of natural and weather conditions present at the time, static and dynamic characteristics of the vessel influencing its manoeuvrability and evaluation of the environmental and infrastructural constraints of the Port.
A range of physiological data will be collected during the experiments, both in the simulated environment and on board ships.
Those recording will include:

1. EMG (Electro Miography),
2. EEG (Electro Encephalogram),
3. ECG (Electro cardiogram)

    (by the mean of electrodes applied on the skin in the chest and the scalp region),

- Respiration amplitude and rate (by the mean of elastic bands around the chest and abdominal region),
- Eye movement and fixation, pupil dilatation (by the mean of an eye tracker device),
- Functional near-infrared spectroscopy (fNIRS).

Each experiment /maneuver is expected to last approximately 30 minutes. You will be kindly asked to participate to 4 simulated experiments in the simulator plus 2 sessions of data collection at resting conditions (1 before and 1 after the experiments). The experiments will be randomly ordered. You will be expected to join the researcher at the Simulator for 1 day, taking into account the duration of the exercises and the time necessary for equipment preparation.
At the end of each experiment you will be asked to complete a NASA TLX questionnaire for mental workload self assessment. Your participation will be agreed with you and the BMP Ops desk, in order to not interfere with you duties or your other commitments.
Data collection for the research is expected to start in November 2012 and to be concluded by March 2013.

- **Are there any possible benefits from participation in this study?**

An important element of this research will be the exploration of the qualitative and quantitative use made by the Pilots of all the previously mentioned systems, aids and information, in order to provide successive criteria to identify possible improvements to human machine interfaces, communications and operational procedures at different levels and standards.
The research could help answering at important questions such as:
What is the level of safety and resiliency granted nowadays in port operations, through the integration of complex systems such as Pilot – Ships – Tugs – VTS (Port Control) – Port logistics and services?
What is the level of expertise required and what are the criteria and the parameters to be considered in order to maintain proper and reliable safety standards in Port operations?
The participation in this study will certainly benefit your awareness of these issues and it will ultimately contribute to a safer and more efficient working environment.
The final report will be available towards the end of 2014 and it will be sent to all participants via e-mail.

- **Are there any possible risks from participation in this study?**

There are no foreseen risks as the required tasks will be carried out in a simulated environment or during routinely operations. Experiments in the simulator will reproduce operational environments with different levels of complexity, exposing participant to different levels of difficulty and workload. Pilots participating to the experiments will be immerged into feasible and realistic scenarios that may happen in real ship handling operations. A component of stress is expected to be experienced by participants in the same amount as it may be experienced in reality.
All the results obtained from the analysis of the data gathered, will be provided and published in a cumulative form, pertaining to a group, remaining completely anonymous in terms of individual identities and related results. The focus of the research is NOT on providing individual professional assessment. NO individual result in any identifiable form, will be provided to BMP, Marine Safety Queensland or any other Company or Agency.
Audio and video recordings, physiological data are functional to exercises' measurement and analysis. Data collected at Smartship and on-board ships will be shared with the AMC researchers for further analysis throughout the PhD programme, only if authorized to do so by participants and always in an anonymous form.
All findings will be securely stored in The Australian Maritime College at the University of Tasmania for 5 years. The computer files will be kept for comparison with future studies of crew performance. No video or audio recording will be published unless directly and specifically authorized, or deliberately masked in order to deceive individual identity.
Data included in the publication of the research will be in a complete anonymous form. Any label or individual identifier will be permanently removed, so that, by no mean, a specific individual can be identified. Stored data will be able to be linked with other data so it can be known that they are about the same subject, but the person's identity will remain completely unknown.

During participation, if you wish to do so, you may discontinue at any time, without providing any explanation.

- **What if I have questions about this research?**

If you would like to discuss any aspect of this study, please feel free to contact Dr Ben Brooks on ph: +61 (0)3-6324-9637 or b.brooks@amc.edu.au. You may also request to be informed when results are available. You are welcome to contact us at that time to discuss any issue relating to the research study.

This study has been approved by the Tasmanian Social Science Human Research Ethics Committee. If you have concerns or complaints about the conduct of this study should contact the Executive Officer of the HREC (Tasmania) Network on +61 (0)3 6226 7479 or email human.ethics@utas.edu.au. The Executive Officer is the person nominated to receive complaints from research participants. You will need to quote the Ethics ref: H12558.

Thank you for taking the time to consider this study.
If you wish to take part in it, please sign the attached consent form.
This information sheet is for you to keep.

Contact Details:

Luca Orlandi
Australian Maritime College
University of Tasmania
M +61
lucaorlandi74@gmail.com
luca.orlandi@utas.edu.au

Dr. Benjamin Brooks (Chief investigator)
Australian Maritime College
University of Tasmania
Email: b.brooks@amc.edu.au

# APPENDIX 3: CONSENT FORM

CONSENT FORM – Participant
Title of PhD research: SEA - Seafarer Expertise Assessment - – Expertise in Maritime Pilotage

1. I have read and understood the 'Information Sheet' for this project.
2. The nature and possible effects of the study have been explained to me.
3. I understand that the study involves audio and videotaped observations of ship mooring operations. I understand that Audio and Video Recordings will be taken using fixed cameras and microphones present in the simulator bridge and exploiting camera and microphone installed on the eye tracker device.
4. I understand that Audio and Video recordings will be not shown to the public or published unless expressly authorized by the participants directly involved in the recordings. I understand that those recording will be exclusively analysed by researchers to obtain anonymous data formatted in a table form.
5. I understand that a structured interview will be conducted. The aim will be to obtain general information regarding years of experience, previous background, motivation for the profession, type of studies and pertinent qualifications. For each of the mentioned elements, open questions (what can you tell me about..) will be posed into a questionnaire that will be handed out to me, just once, at the beginning of the research. The interview will be audio video recorded.
6. I understand that debriefings will be conducted after the experiments allowing me to provide any desired feedback about the exercise just completed or require any explanation or clarification from the researcher. At that stage I will be asked to complete a NASA TLX questionnaire for mental workload self assessment, for which I have received clear information about its meaning and use.
7. I understand that the study involves the continuous unobtrusive recording of physiological parameters during the experiments, which are EEG, ECG, respiration amplitude and rate, eye movement and fixation, pupil dilatation, functional near-infrared spectroscopy (fNIRS).
8. I understand that the data from my observations will be analysed by researchers to identify how performance can be influenced by different factors and working conditions.
9. I understand that the data collected at Smartship Simulator and on board will be shared with the AMC researchers for further analysis throughout the PhD programme and that all data collected will be securely stored in the Australian Maritime College at the University of Tasmania for 5 years.
10. I understand that the computer files will be kept for 5 years, after the conclusion of the research.

    I □ do / □ don’t (please tick the desired option) give consent to reuse my data, in an anonymous form, for comparison with future studies of seafarers assessment.
11. I understand that all the results at the end of the research will be provided referred to groups and never referred to a single individual.

12. I understand that once data collection will be completed, before any form of publication, any label or individual identifier will be permanently removed, so that by no mean, a specific individual can be identified. Data so stored will be able to be linked with other data so it can be known that they are about the same data subject, but the person's identity will remain unknown.
13. I then agree that research data gathered from me for the study may be published provided that I cannot be identified as a participant.
14. I understand that the focus of the research is not on the assessment of my personal technical or non-technical skills and no potential risks are identified for my participation. NO individual results in any identifiable form, will be provided to BMP, Marine Safety Queensland or any other Company or Agency.
15. I understand that I will be able to access a report towards the end of 2014 and that it will be sent to me via email.
16. Any questions that I have asked have been answered to my satisfaction.
17. I agree to participate in this investigation and understand that I may withdraw at any time.

Name of Participant:

Signature: Date:

**Statement by Investigator**

☐ I have explained the project & the implications of participation in it to this volunteer and I believe that the consent is informed and that he/she understands the implications of participation

☐ **The participant has received the Information Sheet where my details have been provided so participants have the opportunity to contact me prior to consenting to participate in this project.**

Name of Investigator Luca Orlandi

Signature of Investigator

# APPENDIX 4: LIST OF QUESTIONS RAISED BY PILOTS

**Communications**

Port control gives updates on weather and traffic movements

Port control has the authority to direct ship movements

What Reporting VHF calls need to be made to advise port authority of entry and where must they be made.

What VHF channels are used in the port.

**Port Characteristics**

Accuracy of Soundings

Anchorage positions

Berth - Bollard Sharing

Berth - Distance between Bollards

Berth - Strongest Lateral Force that can be absorbed or Max Landing Speed

Berth construction and fendering type.

Berth dimensions

Bridge marker presence at the Berth

Depths of water in channel, swing basin and at Berth.

Dimensions of channel and swing basins.

Location and direction of leads

Navigational Aids and Lights Correctly working

Other Ships at Berth

Pipelines or other bottom obstructions that would prevent dredging an anchor

Port Soundings

Port type of bottom

Position of any shore obstructions such as cranes

Simulator Charts - Distance Scales

What are the latest soundings for entry channels.

What is the name of the Berth and the side to alongside.

**Port Regulations**

Local pilotage rules such as passing and overtaking arrangements.

Location of the pilot boarding ground and how far from the fairway is it situated

Min Swinging Distance

Nearest swing basin

Port Speed Limits / Reduction Areas

Temporary Safety Notices to Mariners in place for port.

What are the local port rules for entry

Where are the relevant abort points, emergency anchorage positions, and any speed reduction areas.

**Port Services**

Availability of lines launch/es and linesmen ashore.

Line launch Characteristics

Security Vessel

Sharing linesmen with other ships

Sharing tugs with other ships

Tug Availability

Tug Bollard Pull
Tug Crew understanding of English
Tug Displacement
Tug efficiency and state
Tug size
Tug Type
Using of Ship/Tug lines for making fast tugs
What equipment has been provided to the pilot by their company; e.g. laptop, lifejacket etc.
What is the state of fitness of Tug crew with respect to fatigue & alcohol

**Ship Bridge Equipment**

AIS /PPU Socket availability
Bridge equipment available for providing navigation information.
Bridge equipment location
Bridge equipment. Associated errors.
Prior Arrival - Gyro Check
RPM - Rudder Indicators
Ship AIS System
Ship Bridge Radars
Ship Log
Ship Positioning System
Steering methods

**Ship Characteristics**

Actual Ship Defects, Deficiencies and Limitation
Additional information relating to ships manoeuvring characteristics
Age of the Ship
Availability and readiness of both anchors
Blind Sector Fwd
General state of the ship (painting, hull and superstructure, fitting, hatches, decks...)
Manoeuvrability diagram
Marine Incidents reported for the vessel in the last 12 months
Mooring winches fitted
Notice required by engine-room for manoeuvring and preparing other machinery such as thrusters.
Number of mooring lines
Number of Starts of the Engine
Number of visits that the Ship has already done in the port
Obstructions to Ship bridge visibility
Pilot ladder rigged as per pilot launch request
Prior Arrival - Engine Check
Prior Arrival - Rudder Check
Prior Arrival - Thruster Check
Ship Air Draft
Ship Anchor Cable Length and Shackles
Ship Anchor Type
Ship Anchor Weight

Ship answer on VHF
Ship Beam
Ship Bridge fwd or aft
Ship Bridge or Engine Room Controls
Ship Bridge to Bow
Ship Bridge to Stern
Ship capability to provide a lee
Ship Crash Stop Diagrams
Ship Displacement
Ship Drafts
Ship Engine Timing
Ship Engine Type
Ship Freeboard at Midship
Ship Hull Type
Ship Length Over All
Ship Loading Conditions
Ship Number of Propellers
Ship on time at arrival
Ship particulars
Ship Propeller - Direction of Rotation
Ship Propeller - Type
Ship Rudder angle of neutral effect
Ship Rudder type
Ship Shaft power Ahead
Ship Shaft power Astern
Ship Speed Table
Ship suitability for Tug use
Ship Superstructures and Cranes
Ship Thrusters
Ship Type
Ship UKC
Ship Windage
Ship's Bridge (if at night) light state
Ship's internal accommodation and Bridge smell, tidiness.
Squat Table
SWL of bollards and fair leads for making tugs fast
Type of mooring lines
What is known about the defect/incident history of the Ship from prior visits and other ports.

**Ship Crew**

Bridge understanding of roles and responsibilities as per Passage Plan including: Lookout/ARPA/radar use/courses on ship’s charts/plot position and 7 cables’ notice/question if in doubt/monitor rudder/hand over to next watch
Crew first Impression and general behaviour
Crew Proficiency in English
Crew uniform state

Establish whether the bridge team are happy for pilot to take the con and if Master is to have con for Berthing.

Is there any personnel issue regarding the crew

Need for additional precautions such as extra lookout due to restricted visibility, traffic density, including Ship structures, poor radar performance.

Notice required by crew for manning anchors, taking tugs and Berthing.

Reliability of measures from stern and bow

What is the nationality of the officers and crew and the manning level.

What is the state of fitness of the officers and captain for piloting with respect to fatigue & alcohol

What preparation has been done by Captain and ships officers for pilotage and what are their expectations of pilot.

**Ship Documents**

Check for full understanding and acceptance of the pilots planned courses and speeds for the passage.

Has the ship been issued radio pratique and is she cleared by customs and quarantine for port entry.

Latest charts with corrections

Pilot card

What are the ships company requirements in terms of minimum ukc, speeds and so on.

What are the ships company requirements in terms of tug use.

**Traffic**

Cargo/bunkering ops nearby

Expected manoeuvres of other ships, traffic in port.

What is the other traffic expected, time and order of boarding

**Weather and Environment**

Air Temperature

Current relevant locations with any expected changes

Current speed and direction

DUKC calculation available

Force acting on ship for current (calculation)

How much time has been allocated for the pilotage and what is the scheduled Berthing eta.

How much time is available for Master Pilot exchange

Ice Presence

Sea Temperature

Tide Height

Time of day or night

Waterborne Obstructions

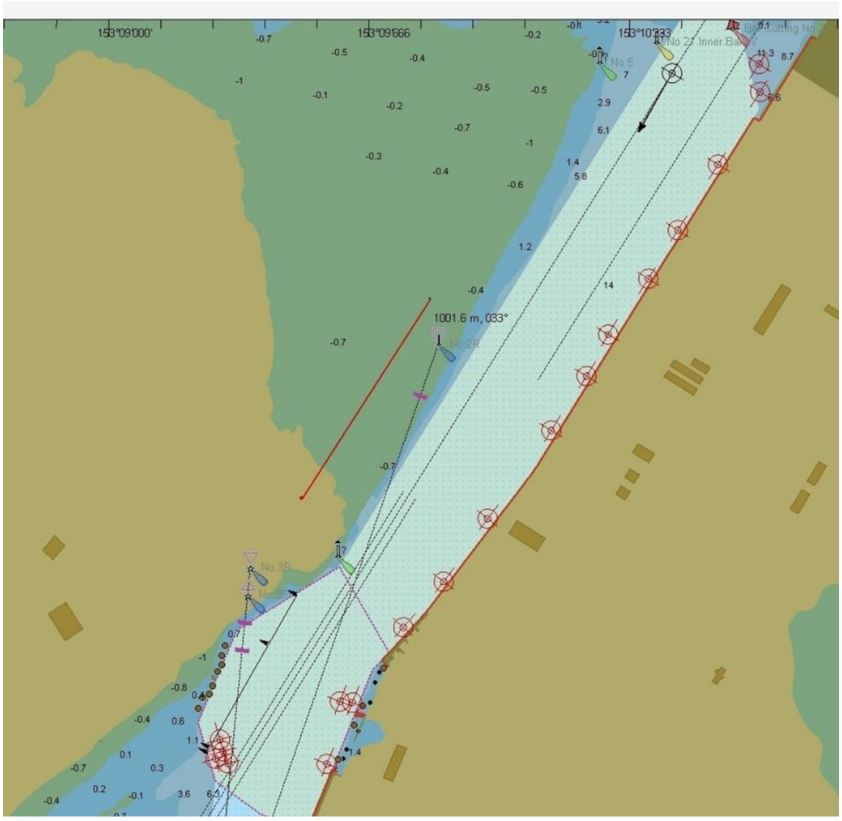
Weather conditions

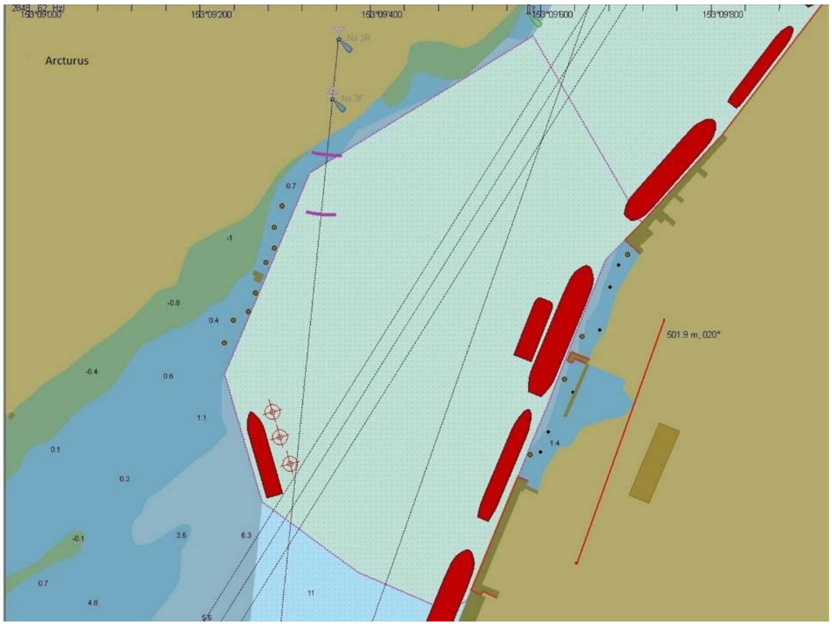
Weather Forecast

Wind local effects from surrounding building and landscape

Wind speed and direction

# APPENDIX 5: PORTS CHARTS



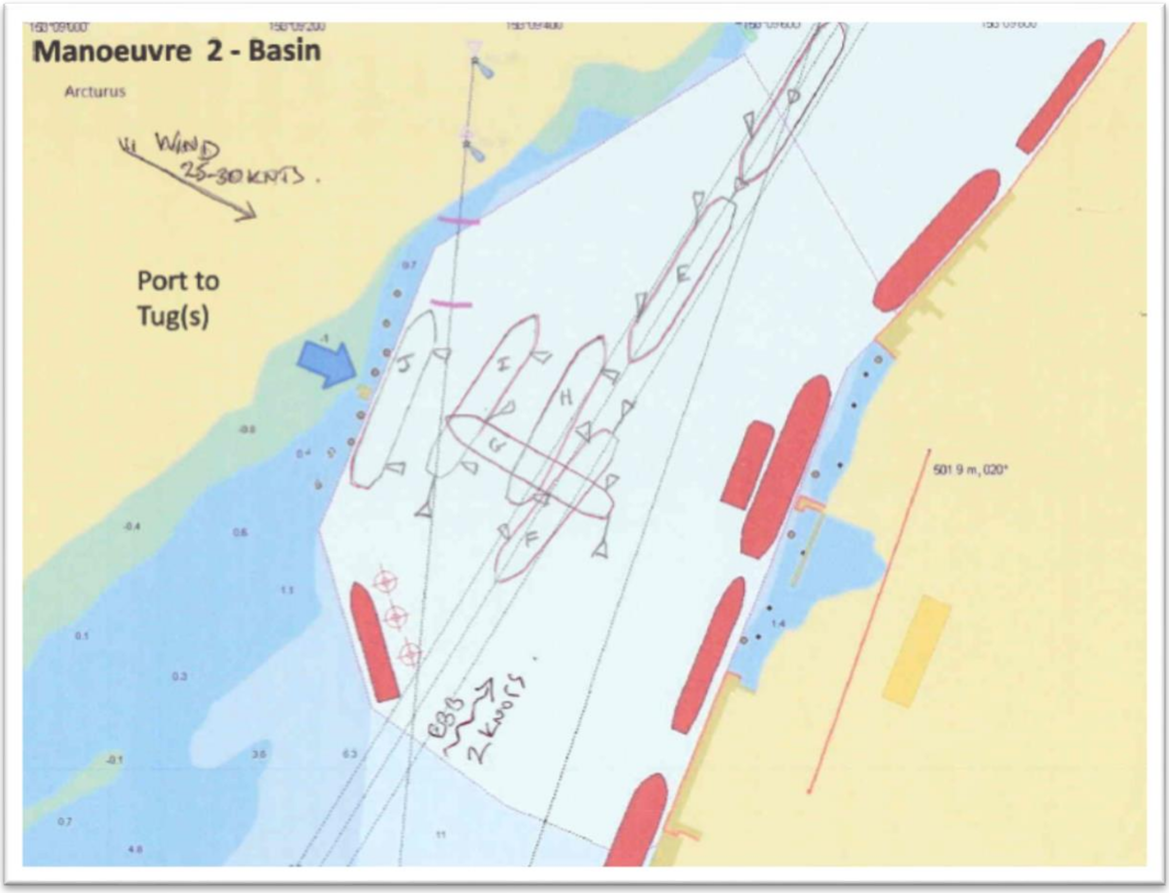




499.9 m, 000°


O - 3485002
M/V Torm Laura (DEIF)
T - 3527001
Queen Mary 2 (DFCH)
500.2 m, 180°
T - 3316001
Burgas (DDBG)
T - 3200001
Ile de Re (DCAA)
T - 3244001
Nordtramp (DCEE)
T - 3434001
MSC Prestige (DEDE)
T - 3598001
Videra (DFJI)

# Appendix 6: Example of completed Detailed Manoeuvring Plan and Chart

| SimTime Secs | Phase | WP Number | Name | Hdg (°T) | SOG Keel (knts) | SOG Side (knts) | Engine Port (%) | Tug1 Pos | Tug1 Power (%) | Tug2 Pos | Tug2 Power (%) | Tug3 Pos | Tug3 Power (%) | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | Approach | 0.5 | Sim-Start | | 6.00 | 0.00 | 50% | StSh | 0% | StQt | 0% | Stern Lea | 0% | 0426 - 134747 |
| 16 | Approach | 1 | Initial | 212 | 6.00 | 0.00 | 50% | StSh | 0% | StQt | 0% | Stern Lea | 0% | 3 Tugs made fast + Tug4 Standby ready to push stbd qtr, no line passed. Ship on the central lead |
| 350 | Approach | 2 | | 212 | 5.00 | 0.00 | 50% | StSh | 0% | StQt | 0% | Laid back | -50% | Reduction of speed expected down to 5 knots, Ship on the central |
| 610 | Approach | 3 | | 212 | 4.00 | 0.00 | 30% | StSh | -50% | StQt | 0% | Laid back | -50% | Reduction of speed expected down to 4 knots |
| 1130 | Approach | 4 | Luggage Point | 212 | 3.00 | 0.00 | 30% | StSh | -50% | StQt | -50% | Laid back | -50% | Reduction of speed expected down to 3 knots. Altering Course to port |
| 1535 | Approach | 5 | | 205 | 2.00 | 0.00 | 15% | StSh | -50% | StQt | -50% | Laid back | -50% | Reduction to 2 knots. |
| 1630 | Approach | 6 | | 205 | 0.00 | 0.00 | 0% | StSh | 0% | StQt | 0% | Laid back | 0% | Shipped stopper over the ground, stemming the current, ready to start the swing bow to stbd. |
| 1631 | Swing | 6.5 | Sim-Swing | | 1.00 | 0.50 | -100% | StSh | 0% | StQt | 75% | Laid back | 0% | Ship in the swing. Expected strong set to stbd and headway. To be counteracted with strong orders on engine and tugs. |
| 1816 | Swing | 7 | | 260 | 1.00 | 0.50 | -100% | StSh | 0% | StQt | 75% | Laid back | 0% | Ship in the swing. Expected strong set to stbd and headway. To be counteracted with strong orders on engine and tugs. |
| 1830 | Swing | 8 | | 300 | 0.70 | 1.00 | -50% | StSh | 0% | StQt | 75% | CLA ast | 0% | Ship swinging bow to stbd |
| 1965 | Swing | 9 | | 330 | 0.50 | 0.50 | -30% | StSh | 50% | StQt | 75% | CLA ast | 0% | Tug1 ready to stop the swing to stbd (risk of overshooting) |
| 2145 | Closing | 9.5 | Sim-Closing | 350 | 0.25 | -0.20 | -15% | StSh | 50% | StQt | 50% | CLA ast | 0% | |
| 2146 | Closing | 10 | | 350 | 0.25 | -0.20 | -15% | StSh | 50% | StQt | 50% | CLA ast | 0% | Ships almost still over the ground, almost parallel to the current. Tugs pushing sideways towards the pier |
| 2696 | ManStop | 10.5 | Sim-Stop | | 0.08 | -0.25 | | | | | | | | |
| | ManStop | 11 | | 15 | 0.00 | 0.00 | -15% | StSh | 25% | StQt | 0% | CLA ast | 0% | Ship along side |


Manoeuvre 2 - Basin
Arcturus
WIND
25-30 KNTS.
Port to
Tug(s)
J
I
H
G
F
E
EBB
2 KNOTS
501.9 m, 020°

# APPENDIX 7: MANOEUVRES CHECK LIST

(1) Check Trackers Memory / Battery status
Save new .evi and do Trackers calibration

(2) Start ECG (box)
Start EEG
Start ECG (computer)

**(3) Sync EEG ECG**

(4) Wait 5-10 min

(5) Check Exercise status
Check Nacos configuration
Check Nacos setup and take over controls
Check countdown clock

(6) Start Trackers (box)
Start Trackers (computer)

(7) Check Trackers Calibration

**(8) Sync EEG ECG**

**(9) Start Exercise**

(10) Sync Sim Clock Trackers

(11) Check EEG
Check ECG
Check Trackers (computer)
Check Trackers (box)

(12) Sync Sim Clock Trackers

**(13) Stop and Save Exercise**

**(14) Sync EEG ECG**

(15) Stop Trackers (box)
Stop Trackers (computer)

(16) Wait 5-10 min
(transfer Trackers SD data)

**(17) Sync EEG ECG**

(18) Stop EEG
Stop ECG (computer)
Stop and Save ECG (box)

(19) Check and Save physio files

(20) Complete Post Ex Forms

(21) Record Debriefing

## APPENDIX 8: SELF ASSESSMENT LIKERT SCALE

| **Level** | **State** | **Description** |
|---|---|---|
| 7 | Extremely Demanding | An extremely demanding situation that it is just about to be out of hand |
| 6 | Very Demanding | A challenging situation that requires the complete attention of the shiphandler, working at almost 100% of his capabilities |
| 5 | Demanding | A situation requiring more attention than normal, but not felt as critical as level 6 |
| 4 | Average | A situation with normal level of involvement where the shiphandler can feel perfectly capable to achieve the desired outcome with a necessary but comfortable level of effort (routine operation) |
| 3 | Easy | An easy situation offering no specific challenge, with required effort below the average |
| 2 | Very Easy | A very comfortable, almost effortless situation |
| 1 | Extremely Easy | A situation of “boredom”, with very little or no involvement at all |

# APPENDIX 9: DESCRIPTION OF BEHAVIOURAL VARIABLES

| BM Name | Targeted behaviours |
|---|---|
| Visual Position Check | A first glance in a certain relative direction (90 ° sector) and then, within 3 seconds, a second glance to a direction perpendicular to the previous. |
| Multiple Position Check | A complete Visual Position Check, followed or anticipated by a glance on a position equipment (Radar screen, chart plotter, Pilot Portable Unit) |
| Visual Direction Check | Shift of gaze from the bow to a direction within 30° off the bow. |
| Multiple Rotation Check | A complete Visual Direction Check, followed or anticipated by a glance on a rate of turn sensor (ROT indicator) |
| Visual Speed Check | A glance at the beam of the vessel, followed or anticipated by a glance on a speed sensor (LOG or GPS SOG indicator) |
| Plan Check | Glance at the plan |
| Pilot Orders | Pilot order to bridge team member or tug (via radio) |

# APPENDIX 10: GENERAL INTERVIEW

**SEA Seafarer Expertise Assessment**
**Expertise in Maritime Pilotage**
**PhD Research**

## General Interview

Pilot _______________ Date of Interview _______________ Year Of Birth _______________

Spectacles _______________ Type of License _______________ Year as BMP ______________

| |
|---|
| **What is you fatigue state today (briefly describe anything of interest happened in the last 24h that could influence the state of fatigue)** |
| ____________________________________________ |
| **Are you taking Medications (substances that may interact with physiological measures)** |
| ____________________________________________ |
| **Type of Glasses / Correction used when piloting (also reading glasses or contact lenses)** |
| ____________________________________________ |
| **Medical conditions that may interfere with Manoeuvres (if any)** |
| ____________________________________________ |
| **Caffeine Nicotine or other substances assumed recently (if any)** |
| ____________________________________________ |
| **Previous experience with similar ships to those used in the manoeuvres(if any)** |
| ____________________________________________ |
| **When did you start to go at sea and why (briefly describe the initial motivation to start a career at sea)** |
| ____________________________________________ |

| |
|---|
| ____________________________________________________________________________ |
| **What are the studies that you did, relevant to your Profession (briefly describe any study or professional course or certification relevant to the Profession)** |
| ____________________________________________________________________________ |
| **What is your relevant experience on board ships (briefly describe the type of ships you have been on board and their peculiar characteristics in terms of manoeuvrability, limitations...)** |
| ____________________________________________________________________________ |
| **What is your relevant experience as a Pilot (briefly describe the type of ships and Ports you have been working and their peculiar characteristics in terms of manoeuvrability, limitations...)** |
| ____________________________________________________________________________ |
| **What do you enjoy most in your profession (briefly describe the aspects most appreciated, actual motivation, positive drives)** |
| ____________________________________________________________________________ |
| **What don't you like about your profession (briefly describe the aspects less appreciated, negative drives)** |
| ____________________________________________________________________________ |
| **Is there any particular activity you engage that you use to improve your skills as a Pilot, out of the normal job requirements (briefly describe if there is any particular deliberate practice, study, exercise used to improve outcomes)** |
| ____________________________________________________________________________ |
| **Do you practise sport (-)** |
| ____________________________________________________________________________ |

## APPENDIX 11: REMINDER EMAIL BEFORE SIMULATOR SESSION

From: Luca Orlandi
Sent:
To:
Subject: Day at the Simulator

Dear Capt. ____________________________

Many thanks for the documents all well received!
Looking forward to see you tomorrow!
I will be ready from 09.00 but fell free to take your time, if the traffic then could be an issue and you prefer to join me later.

Just few things to remember!

1) If you like to have any coffee in the morning, please have it at least an hour before coming to the Simulator and not later, as you well imagine, it could interfere with some measures we are going to take!)
2) If you smoke, consider that nicotine will interfere up to 45 minutes after you had your cigarette.
3) Be aware that you will be wearing some eye tracking glasses that can go over the glasses you may normally use, but once put on, they will have to stay for the entire duration of each exercise! This could be an issue if you have to use "reading glasses" that you wear only at times. Please let me know if you have any issue with that!
4) Feel free to come with your PPU, if you would like to use it. We will try to interface it with the Sim.
5) In case you may have other commitments for the day, you may want to know that the entire duration of your staying at the Simulator could be around 9-10 hours all together. From 0900 until 1800 (possibly later..) depending on how we proceed. It would be very nice, if time could not be a big issue for that day, allowing us a bit of flexibility.
6) Since it's going to be a long day, I would suggest to carry with you some "comfort food" (fruit or snacks) that you may want to bite from time to time! There is a mess/bar just crossing the street, where we can have a sandwich or a burger at lunchtime.

Feel free to contact me at any time, should any further information be required.
Really many thanks again and looking forward to seeing you tomorrow!
Regards

Luca

## APPENDIX 12: LIST OF BEHAVIOURAL MARKERS USED TO CODE VIDEO RECORDINGS

| Code | Group |
|---|---|
| ShpPsnChkVFS | Sequences |
| ShpPsnChkVSF | Sequences |
| ShpPsnChkVSA | Sequences |
| ShpPsnChkVAS | Sequences |
| ShpPsnChkVAP | Sequences |
| ShpPsnChkVPA | Sequences |
| ShpPsnChkVPF | Sequences |
| ShpPsnChkVFP | Sequences |
| ShpPsnChkV* | Sequences |
| ShpPsnChkTVM | Sequences |
| ShpPsnChkTMV | Sequences |
| ShpPsnChkT* | Sequences |
| ShpPsnChk* | Sequences |
| ShpDirChkVFI | Sequences |
| ShpDirChkVIF | Sequences |
| ShpDirChkVBF | Sequences |
| ShpDirChkVFB | Sequences |
| ShpDirChkV* | Sequences |
| ShpDirChkTVM | Sequences |
| ShpDirChkTMV | Sequences |
| ShpDirChkT* | Sequences |
| ShpDirChk* | Sequences |
| ShpSpdChkVIF | Sequences |
| ShpSpdChkVFI | Sequences |
| ShpSpdChkV* | Sequences |
| ShpSpdChkTVM | Sequences |
| ShpSpdChkTMV | Sequences |
| ShpSpdChkT* | Sequences |
| ShpSpdChk* | Sequences |
| OrdRddrChkOIC | Sequences |
| OrdRddrChkOCI | Sequences |
| OrdRddrChk* | Sequences |
| OrdEngChkOIC | Sequences |
| OrdEngChkOCI | Sequences |
| OrdEngChk* | Sequences |
| OrdThrChkOIC | Sequences |
| OrdThrChkOCI | Sequences |
| OrdThrChk* | Sequences |
| OrdTugChkOVC | Sequences |
| OrdTugChkOCV | Sequences |
| OrdTugChk* | Sequences |
| Ord*Chk | Sequences |
| EqmtPsnChk01 | Sequences |
| EqmtPsnChk02 | Sequences |

| Code | Group |
|---|---|
| Hdg* | Vis_Sensor |
| Spd* | Vis_Sensor |
| Current | Vis_Sensor |
| EngInd | Vis_Sensor |
| HdgOvr | Vis_Sensor |
| HdgSns | Vis_Sensor |
| Nacos | Vis_Sensor |
| PPU | Vis_Sensor |
| PsnSns | Vis_Sensor |
| Qastor | Vis_Sensor |
| Radio | Vis_Sensor |
| RddrInd | Vis_Sensor |
| ROTInd | Vis_Sensor |
| SpdGPSSns | Vis_Sensor |
| SpdLOGOvr | Vis_Sensor |
| SpdLOGSns | Vis_Sensor |
| ThrstrInd | Vis_Sensor |
| TrafRDRAIS | Vis_Sensor |
| UKC | Vis_Sensor |
| Wind | Vis_Sensor |
| BowExt | Vis_ExtCue |
| Document | Vis_ExtCue |
| NavAids | Vis_ExtCue |
| Plan | Vis_ExtCue |
| FixObject | Vis_ExtCue |
| ScanObject | Vis_ExtCue |
| TgtDistr | Vis_ExtCue |
| TgtInter | Vis_ExtCue |
| TgtUnexp | Vis_ExtCue |
| TgtVTS | Vis_ExtCue |
| ThrstrFlush | Vis_ExtCue |
| TrafExt | Vis_ExtCue |
| TugExt | Vis_ExtCue |
| EngAct | Vis_Actuator |
| RuddrAct | Vis_Actuator |
| ThrstrAct | Vis_Actuator |
| Pilot_CLP* | Voi_Pilot |
| Pilot_CLP-Brdg | Voi_Pilot |
| Pilot_CLP-Tug | Voi_Pilot |
| Pilot_AKN* | Voi_Pilot |
| Pilot_NAK* | Voi_Pilot |
| Pilot_Request | Voi_Pilot |
| Pilot_Order | Voi_Pilot |
| Brg*Ord | Voi_PilDetail |

| Code | Group |
|---|---|
| EqmtPsnChk* | Sequences |
| EqmtHdgChk01 | Sequences |
| EqmtHdgChk02 | Sequences |
| EqmtHdgChk* | Sequences |
| EqmtSpdChk01 | Sequences |
| EqmtSpdChk02 | Sequences |
| EqmtSpdChk* | Sequences |
| Eqmt???Chk* | Sequences |
| EnvChk01* | Sequences |
| ShpROTChkVBO | Sequences |
| ShpROTChkVOB | Sequences |
| ShpROTChkTVI | Sequences |
| ShpROTChkTIV | Sequences |
| ShpROTChkV* | Sequences |
| ShpROTChkT* | Sequences |
| Bridge_Info | Voi_Other |
| Bridge_Order | Voi_Other |
| Bridge_Request | Voi_Other |
| Tug_Info | Voi_Other |
| Tug_Request | Voi_Other |
| VTS_Info | Voi_Other |
| VTS_Request | Voi_Other |
| | |
| | |
| | |

| Code | Group |
|---|---|
| BrgAnchOrd | Voi_PilDetail |
| BrgEngOrd | Voi_PilDetail |
| BrgHdgOrd | Voi_PilDetail |
| BrgROTOrd | Voi_PilDetail |
| BrgRuddrOrd | Voi_PilDetail |
| BrgThrBowOrd | Voi_PilDetail |
| BrgWingTrfOrd | Voi_PilDetail |
| Tug?Ord | Voi_PilDetail |
| Tug1Ord | Voi_PilDetail |
| Tug2Ord | Voi_PilDetail |
| Tug3Ord | Voi_PilDetail |
| Tug4Ord | Voi_PilDetail |
| Pilot_Comm | Voi_Pilot |
| BrgMPSCom | Voi_PilDetail |
| ComntCom | Voi_PilDetail |
| VHF*Com | Voi_PilDetail |
| VHFGenTrafCom | Voi_PilDetail |
| VHFLinesCom | Voi_PilDetail |
| VHFRptPntCom | Voi_PilDetail |
| VHFTrgtDistrCom | Voi_PilDetail |
| VHFTrgtInterCom | Voi_PilDetail |
| VHFTrgtUnexpCom | Voi_PilDetail |
| VHFTrgtVTSCom | Voi_PilDetail |
| VHFTugsCom | Voi_PilDetail |
| VHFVTSCom | Voi_PilDetail |

# APPENDIX 13: EXAMPLE OF VIDEO / AUDIO CODING FILE

| Entry | Exit | Vis_Sensor | Vis_ExtCue | Vis_Actuator | Notes | Spare | Voi_Pilot | Voi_PilDetail | Voi_Other |
|---|---|---|---|---|---|---|---|---|---|
| 00:01:19:08 | 00:01:19:09 | | | | 10:00:15 SimTime - 208930 EyeTrk Frame - 2376 Coding Frame | SynStart | | | |
| 00:01:23:29 | 00:01:25:10 | | | RuddrAct | | | | | |
| 00:01:25:10 | 00:01:29:15 | | | | | NoCode | | | |
| 00:01:25:29 | 00:01:27:20 | | | | | | Pilot_Order | BrgEngOrd-Ahead-OneQtr | |
| 00:01:27:20 | 00:01:28:19 | | | | | | | | Bridge_AKN |
| 00:01:28:19 | 00:01:30:25 | | | | 212 | | Pilot_Order | BrgHdgOrd | |
| 00:01:29:15 | 00:01:31:26 | HdgSns | | | | | | | |
| 00:01:30:25 | 00:01:31:21 | | | | | | | | Bridge_AKN |
| 00:01:31:21 | 00:01:32:21 | | | | | | Pilot_AKN-Brdg | | |
| 00:01:31:26 | 00:01:32:18 | | BowExt_000 | | | | | | |
| 00:01:32:18 | 00:01:33:01 | | TugExt | | | | | | |
| 00:01:33:01 | 00:01:38:17 | Radio | | | | | | | |
| 00:01:38:03 | 00:01:39:18 | | | | Radio Check | | Pilot_Comm | VHFTugsCom | |
| 00:01:38:17 | 00:01:38:29 | SpdLOGSns | | | | | | | |
| 00:01:38:29 | 00:01:40:17 | | TugExt | | | | | | |
| 00:01:39:18 | 00:01:43:11 | | | | | | | | Tug_AKN |
| 00:01:40:17 | 00:01:41:17 | | BowExt_000 | | | | | | |
| 00:01:41:17 | 00:01:43:05 | | TugExt | | | | | | |
| 00:01:43:05 | 00:01:43:29 | | BowExt_000 | | | | | | |
| 00:01:43:11 | 00:01:44:18 | | | | | | Pilot_CLP-Tug | | |
| 00:01:43:29 | 00:01:46:16 | | TugExt | | | | | | |
| 00:01:46:16 | 00:01:48:14 | | | | | NoCode | | | |
| 00:01:48:14 | 00:01:50:03 | | BowExt_000 | | | | | | |
| 00:01:50:01 | 00:01:56:29 | | | | Interacting Targer spotted | | Pilot_Comm | BrgMPSCom | |
| 00:01:50:03 | 00:01:54:21 | | TgtInter | | | | | | |
| 00:01:54:21 | 00:01:55:21 | | | | | NoCode | | | |
| 00:01:55:21 | 00:01:57:15 | | BowExt_000 | | | | | | |
| 00:01:57:15 | 00:01:59:01 | PsnSns | | | port (main - chart overlay) | | | | |
| 00:01:59:01 | 00:02:00:02 | PsnSns | | | stbd (secondary) | | | | |
| 00:02:00:02 | 00:02:05:24 | | FixObject_000 | | | | | | |
| 00:02:05:24 | 00:02:08:08 | | BowExt_000 | | | | | | |
| 00:02:08:08 | 00:02:10:12 | PsnSns | | | port | | | | |
| 00:02:10:12 | 00:02:12:06 | PsnSns | | | stbd | | | | |
| 00:02:12:06 | 00:02:13:09 | HdgSns | | | | | | | |
| 00:02:13:09 | 00:02:14:14 | PsnSns | | | port | | | | |
| 00:02:14:14 | 00:02:16:01 | HdgSns | | | | | | | |
| 00:02:14:26 | 00:02:19:19 | | | | change scale on Radar | | Pilot_Request | | |
| 00:02:16:01 | 00:02:17:24 | PsnSns | | | stbd | | | | |
| 00:02:17:24 | 00:02:18:22 | Nacos | | | | | | | |
| 00:02:18:22 | 00:02:20:26 | PsnSns | | | stbd | | | | |
| 00:02:19:19 | 00:02:20:28 | | | | Scale Modified | | | | Bridge_Info |

## APPENDIX 14: ADDITIONAL STATISTICS (NOT PUBLISHED)

## Paper III - Measuring mental workload and physiological reactions in marine pilots: building bridges towards redlines of performance.

### NASA TLX

**Tests of Normality**

| Port | | Diff | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|---|---|
| | | | Statistic | df | Sig. | Statistic | df | Sig. |
| 0 | NASA TLX | 0 | .159 | 10 | .200* | .972 | 10 | .911 |
| | | 1 | .251 | 10 | .074 | .884 | 10 | .146 |
| 1 | NASA TLX | 0 | .158 | 10 | .200* | .943 | 10 | .588 |
| | | 1 | .156 | 10 | .200* | .976 | 10 | .941 |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

**Levene's Test of Equality of Error Variances[a]**

Dependent Variable: NASA TLX

| F | df1 | df2 | Sig. |
|---|---|---|---|
| 1.005 | 3 | 36 | .402 |

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Diff + Port + Diff * Port

## Self Assessment Likert Scale

**Tests of Normality**

| Diff | Port | Phase | | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Statistic | df | Sig. | Statistic | df | Sig. |
| 0 | 0 | 1 | Likert Scale | .258 | 10 | .057 | .867 | 10 | .092 |
| | | 2 | Likert Scale | .212 | 10 | .200* | .897 | 10 | .204 |
| | | 3 | Likert Scale | .157 | 10 | .200* | .931 | 10 | .461 |
| | 1 | 1 | Likert Scale | .264 | 10 | .047 | .892 | 10 | .177 |
| | | 2 | Likert Scale | .143 | 10 | .200* | .922 | 10 | .378 |
| | | 3 | Likert Scale | .200 | 10 | .200* | .932 | 10 | .468 |
| 1 | 0 | 1 | Likert Scale | .147 | 10 | .200* | .972 | 10 | .908 |
| | | 2 | Likert Scale | .161 | 9 | .200* | .920 | 9 | .392 |
| | | 3 | Likert Scale | .157 | 9 | .200* | .923 | 9 | .422 |
| | 1 | 1 | Likert Scale | .146 | 10 | .200* | .965 | 10 | .844 |
| | | 2 | Likert Scale | .231 | 9 | .182 | .939 | 9 | .575 |
| | | 3 | Likert Scale | .204 | 7 | .200* | .912 | 7 | .408 |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

**Levene's Test of Equality of Error Variances[a]**

Dependent Variable: Likert Scale

| F | df1 | df2 | Sig. |
|---|---|---|---|
| 1.119 | 11 | 102 | .354 |

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Diff + Port + Phase + Diff * Port + Diff * Phase + Port * Phase + Diff * Port * Phase

## HR

**Tests of Normality**

| Diff | Port | Phase | | Kolmogorov-Smirnov[a] Statistic | df | Sig. | Shapiro-Wilk Statistic | df | Sig. |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | HR | .164 | 10 | .200* | .962 | 10 | .806 |
| | | 1 | HR | .126 | 10 | .200* | .985 | 10 | .987 |
| | | 2 | HR | .214 | 10 | .200* | .844 | 10 | .050 |
| | | 3 | HR | .143 | 10 | .200* | .950 | 10 | .671 |
| | | 4 | HR | .170 | 9 | .200* | .902 | 9 | .266 |
| | 1 | 0 | HR | .206 | 10 | .200* | .898 | 10 | .210 |
| | | 1 | HR | .154 | 10 | .200* | .931 | 10 | .462 |
| | | 2 | HR | .118 | 10 | .200* | .961 | 10 | .797 |
| | | 3 | HR | .179 | 10 | .200* | .935 | 10 | .498 |
| | | 4 | HR | .202 | 9 | .200* | .899 | 9 | .246 |
| 1 | 0 | 0 | HR | .229 | 10 | .146 | .875 | 10 | .115 |
| | | 1 | HR | .210 | 10 | .200* | .913 | 10 | .303 |
| | | 2 | HR | .213 | 8 | .200* | .916 | 8 | .397 |
| | | 3 | HR | .272 | 8 | .083 | .844 | 8 | .084 |
| | | 4 | HR | .215 | 8 | .200* | .919 | 8 | .425 |
| | 1 | 0 | HR | .201 | 9 | .200* | .959 | 9 | .790 |
| | | 1 | HR | .184 | 9 | .200* | .958 | 9 | .782 |
| | | 2 | HR | .257 | 8 | .129 | .844 | 8 | .083 |
| | | 3 | HR | .239 | 6 | .200* | .840 | 6 | .131 |
| | | 4 | HR | .176 | 7 | .200* | .922 | 7 | .483 |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

**Levene's Test of Equality of Error Variances[a]**

Dependent Variable: HR

| F | df1 | df2 | Sig. |
|---|---|---|---|
| 1.170 | 19 | 161 | .289 |

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Diff + Port + Phase + Diff * Port + Diff * Phase + Port * Phase + Diff * Port * Phase

## LF/HF

**Tests of Normality**

| Diff | Port | Phase | | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Statistic | df | Sig. | Statistic | df | Sig. |
| 0 | 0 | 0 | LFHF | .130 | 9 | .200* | .948 | 9 | .673 |
| | | 1 | LFHF | .160 | 10 | .200* | .968 | 10 | .867 |
| | | 2 | LFHF | .163 | 10 | .200* | .947 | 10 | .630 |
| | | 3 | LFHF | .198 | 10 | .200* | .931 | 10 | .460 |
| | | 4 | LFHF | .130 | 9 | .200* | .976 | 9 | .940 |
| | 1 | 0 | LFHF | .184 | 10 | .200* | .943 | 10 | .591 |
| | | 1 | LFHF | .159 | 10 | .200* | .986 | 10 | .989 |
| | | 2 | LFHF | .272 | 10 | .034 | .846 | 10 | .052 |
| | | 3 | LFHF | .140 | 10 | .200* | .981 | 10 | .972 |
| | | 4 | LFHF | .129 | 8 | .200* | .982 | 8 | .973 |
| 1 | 0 | 0 | LFHF | .155 | 10 | .200* | .955 | 10 | .728 |
| | | 1 | LFHF | .247 | 10 | .085 | .783 | 10 | .009 |
| | | 2 | LFHF | .169 | 8 | .200* | .964 | 8 | .844 |
| | | 3 | LFHF | .169 | 8 | .200* | .955 | 8 | .763 |
| | | 4 | LFHF | .228 | 7 | .200* | .878 | 7 | .217 |
| | 1 | 0 | LFHF | .172 | 9 | .200* | .952 | 9 | .715 |
| | | 1 | LFHF | .252 | 9 | .102 | .860 | 9 | .096 |
| | | 2 | LFHF | .161 | 8 | .200* | .986 | 8 | .986 |
| | | 3 | LFHF | .188 | 6 | .200* | .963 | 6 | .839 |
| | | 4 | LFHF | .189 | 7 | .200* | .920 | 7 | .471 |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

**Levene's Test of Equality of Error Variances[a]**

Dependent Variable: LFHF

| F | df1 | df2 | Sig. |
|---|---|---|---|
| 1.294 | 19 | 158 | .194 |

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Diff + Port + Phase + Diff * Port + Diff * Phase + Port * Phase + Diff * Port * Phase

## Pupil Dilation

**Tests of Normality**

| Diff | Port | Phase | | Kolmogorov-Smirnov[a] Statistic | df | Sig. | Shapiro-Wilk Statistic | df | Sig. |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | Pupil | .181 | 10 | .200* | .928 | 10 | .426 |
| | | 2 | Pupil | .129 | 10 | .200* | .967 | 10 | .861 |
| | | 3 | Pupil | .145 | 10 | .200* | .965 | 10 | .843 |
| | 1 | 1 | Pupil | .200 | 10 | .200* | .940 | 10 | .555 |
| | | 2 | Pupil | .332 | 10 | .003 | .779 | 10 | .008 |
| | | 3 | Pupil | .177 | 10 | .200* | .854 | 10 | .064 |
| 1 | 0 | 1 | Pupil | .132 | 10 | .200* | .968 | 10 | .877 |
| | | 2 | Pupil | .194 | 8 | .200* | .899 | 8 | .282 |
| | | 3 | Pupil | .248 | 8 | .158 | .829 | 8 | .058 |
| | 1 | 1 | Pupil | .149 | 10 | .200* | .965 | 10 | .845 |
| | | 2 | Pupil | .198 | 9 | .200* | .955 | 9 | .743 |
| | | 3 | Pupil | .134 | 7 | .200* | .990 | 7 | .993 |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

**Levene's Test of Equality of Error Variances[a]**

Dependent Variable: Pup_Dil_Norm

| F | df1 | df2 | Sig. |
|---|---|---|---|
| 1.784 | 11 | 100 | .067 |

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Diff + Port + Phase + Diff * Port + Diff * Phase + Port * Phase + Diff * Port * Phase

## EEG Beta 1

**Tests of Normality**

| Diff | Port | Phase | | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Statistic | df | Sig. | Statistic | df | Sig. |
| 0 | 0 | 0 | EEG_B1 | .177 | 9 | .200* | .958 | 9 | .777 |
| | | 1 | EEG_B1 | .176 | 9 | .200* | .908 | 9 | .305 |
| | | 2 | EEG_B1 | .227 | 9 | .200* | .953 | 9 | .726 |
| | | 3 | EEG_B1 | .199 | 9 | .200* | .920 | 9 | .393 |
| | | 4 | EEG_B1 | .265 | 8 | .103 | .896 | 8 | .268 |
| | 1 | 0 | EEG_B1 | .318 | 9 | .009 | .789 | 9 | .015 |
| | | 1 | EEG_B1 | .277 | 9 | .045 | .900 | 9 | .251 |
| | | 2 | EEG_B1 | .215 | 9 | .200* | .943 | 9 | .611 |
| | | 3 | EEG_B1 | .115 | 9 | .200* | .991 | 9 | .997 |
| | | 4 | EEG_B1 | .243 | 7 | .200* | .865 | 7 | .168 |
| 1 | 0 | 0 | EEG_B1 | .131 | 10 | .200* | .960 | 10 | .782 |
| | | 1 | EEG_B1 | .164 | 10 | .200* | .973 | 10 | .920 |
| | | 2 | EEG_B1 | .209 | 8 | .200* | .905 | 8 | .323 |
| | | 3 | EEG_B1 | .210 | 8 | .200* | .936 | 8 | .568 |
| | | 4 | EEG_B1 | .149 | 8 | .200* | .957 | 8 | .780 |
| | 1 | 0 | EEG_B1 | .153 | 9 | .200* | .928 | 9 | .459 |
| | | 1 | EEG_B1 | .255 | 9 | .095 | .829 | 9 | .044 |
| | | 2 | EEG_B1 | .197 | 8 | .200* | .971 | 8 | .903 |
| | | 3 | EEG_B1 | .247 | 7 | .200* | .842 | 7 | .103 |
| | | 4 | EEG_B1 | .180 | 8 | .200* | .959 | 8 | .798 |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

**Levene's Test of Equality of Error Variances[a]**

Dependent Variable: EEG_B1

| F | df1 | df2 | Sig. |
|---|---|---|---|
| 1.142 | 19 | 152 | .315 |

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Diff + Port + Phase + Diff * Port + Diff * Phase + Port * Phase + Diff * Port * Phase

## EEG Beta 2

**Tests of Normality**

| Diff | Port | Phase | | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Statistic | df | Sig. | Statistic | df | Sig. |
| 0 | 0 | 0 | EEG_B2 | .156 | 9 | .200* | .937 | 9 | .556 |
| | | 1 | EEG_B2 | .173 | 9 | .200* | .921 | 9 | .403 |
| | | 2 | EEG_B2 | .193 | 9 | .200* | .946 | 9 | .650 |
| | | 3 | EEG_B2 | .248 | 9 | .118 | .866 | 9 | .111 |
| | | 4 | EEG_B2 | .210 | 8 | .200* | .918 | 8 | .410 |
| | 1 | 0 | EEG_B2 | .201 | 9 | .200* | .948 | 9 | .672 |
| | | 1 | EEG_B2 | .217 | 9 | .200* | .935 | 9 | .530 |
| | | 2 | EEG_B2 | .165 | 9 | .200* | .942 | 9 | .605 |
| | | 3 | EEG_B2 | .160 | 9 | .200* | .933 | 9 | .513 |
| | | 4 | EEG_B2 | .191 | 7 | .200* | .890 | 7 | .277 |
| 1 | 0 | 0 | EEG_B2 | .106 | 10 | .200* | .992 | 10 | .998 |
| | | 1 | EEG_B2 | .212 | 10 | .200* | .927 | 10 | .421 |
| | | 2 | EEG_B2 | .241 | 8 | .190 | .862 | 8 | .127 |
| | | 3 | EEG_B2 | .144 | 8 | .200* | .972 | 8 | .912 |
| | | 4 | EEG_B2 | .158 | 8 | .200* | .969 | 8 | .893 |
| | 1 | 0 | EEG_B2 | .212 | 9 | .200* | .872 | 9 | .128 |
| | | 1 | EEG_B2 | .288 | 9 | .030 | .769 | 9 | .009 |
| | | 2 | EEG_B2 | .155 | 8 | .200* | .974 | 8 | .924 |
| | | 3 | EEG_B2 | .238 | 7 | .200* | .843 | 7 | .105 |
| | | 4 | EEG_B2 | .169 | 8 | .200* | .953 | 8 | .745 |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

**Levene's Test of Equality of Error Variances[a]**

Dependent Variable: EEG_B2

| F | df1 | df2 | Sig. |
|---|---|---|---|
| .685 | 19 | 152 | .830 |

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Diff + Port + Phase + Diff * Port + Diff * Phase + Port * Phase + Diff * Port * Phase